

2014

# The effectiveness of linking instruction on NNS speech perception and production

Ghinwa Alameen  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>



Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), and the [Linguistics Commons](#)

---

## Recommended Citation

Alameen, Ghinwa, "The effectiveness of linking instruction on NNS speech perception and production" (2014). *Graduate Theses and Dissertations*. 14004.  
<https://lib.dr.iastate.edu/etd/14004>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

# **The effectiveness of linking instruction on NNS speech perception and production**

by

**Ghinwa Alameen**

A dissertation submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Applied Linguistics and Technology

Program of Study Committee:  
John M. Levis, Major Professor  
Janet Anderson-Hsieh  
Bethany E. Gray  
Donna M. Niday  
Tammy Slater  
Horabail Venkatagiri

Iowa State University

Ames, Iowa

2014

Copyright © Ghinwa Alameen, 2014. All rights reserved.

To Haifa and Marwan

To Syria..

## TABLE OF CONTENTS

<b>LIST OF FIGURES .....</b>	<b>vi</b>
<b>LIST OF TABLES .....</b>	<b>vii</b>
<b>NOMENCLATURE.....</b>	<b>ix</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>x</b>
<b>ABSTRACT.....</b>	<b>ix</b>
<b>CHAPTER I: INTRODUCTION .....</b>	<b>1</b>
Statement of the Problem.....	3
Purpose of the Study .....	4
Research Questions .....	5
Significance of the Study .....	6
Organization of the Dissertation .....	6
<b>CHAPTER II: LITERATURE REVIEW .....</b>	<b>8</b>
Connected Speech Processes .....	8
Connected speech features .....	9
Linking .....	12
Definition of linking .....	12
Types of linking .....	13
Connected Speech Perception.....	15
Speech segmentation.....	16
Effects of CSPs on listening perception and comprehension .....	17
Effectiveness of CSP perceptual training on listening.....	18
Connected Speech Production .....	25
Connected speech in NS and NNS production .....	25
Effectiveness of CSP instruction on L2 production.....	26
The Effect of Word Frequency .....	28
Electronic Visual Feedback .....	30
Types of electronic visual feedback.....	30
Using waveforms in teaching linking .....	34
EVF software .....	36
Advantages and limitations of EVF.....	37
Guidelines for using EVF .....	38
Effectiveness of EVF on L2 pronunciation .....	40
Student perspectives on using EVF .....	42
Chapter Summary .....	43
<b>CHAPTER III: METHODS .....</b>	<b>44</b>
Research Design.....	44
Participants and Setting Description.....	46

Procedures .....	48
Training materials .....	49
Chapter Summary .....	57
<b>CHAPTER IV: THE EFFECTIVENESS OF LINKING INSTRUCTION ON PERCEPTION .....</b>	<b>58</b>
Materials .....	58
Perception tests .....	58
Perception Test Stimuli .....	59
Analysis .....	59
Results .....	61
Research question 1: The effectiveness of training .....	62
Research question 2: The long-term effects of training .....	63
Research question 3: Transfer to novel contexts .....	66
Discussion .....	68
Research question 1: The effectiveness of training .....	68
Research question 2: The long-term effects of training .....	73
Research question 3: Transfer to novel contexts .....	74
Chapter Summary .....	75
<b>CHAPTER V: THE EFFECTIVENESS OF LINKING INSTRUCTION ON PRODUCTION .....</b>	<b>76</b>
Materials .....	76
Production tests .....	76
Production tests stimuli .....	77
Analysis .....	79
Preparing production data .....	79
Rating criteria: Revisiting the definition of linking .....	79
Second rater training .....	80
Production analysis procedure .....	81
Results .....	83
Research question 1: The effectiveness of training .....	83
Research question 2: The long-term effects of training .....	85
Research question 3: Transfer to novel contexts .....	88
Research question 4: The effect of word frequency .....	90
Discussion .....	92
Research question 1: The effectiveness of training .....	92
Research question 2: The long-term effects of training .....	98
Research question 3: Transfer to novel contexts .....	99
Research question 4: The effect of word frequency .....	100
Chapter Summary .....	102
<b>CHAPTER VI: LEARNERS' PERSPECTIVES ON THE EFFECTIVENESS OF LINKING INSTRUCTION .....</b>	<b>103</b>
Post-training Questionnaire .....	103
Analysis .....	105

Results .....	106
Discussion .....	110
Chapter Summary .....	112
<b>CHAPTER VII: CONCLUSION.....</b>	<b>113</b>
Overview of the Study .....	113
Summary of Major Findings .....	114
The effectiveness of training .....	115
The long-term effects of training .....	117
Transfer to novel contexts .....	118
The effect of word frequency .....	118
Learners' perception of the training .....	119
Pedagogical Implications .....	121
Limitations of the Study .....	125
Recommendations for Future Research .....	126
Conclusion .....	128
<b>REFERENCES.....</b>	<b>129</b>
<b>APPENDIX I. PRE-TRAINING QUESTIONNAIRE.....</b>	<b>140</b>
<b>APPENDIX II. INFORMED CONSENT DOCUMENT .....</b>	<b>142</b>
<b>APPENDIX III. AV GROUP TRAINING HANDOUT .....</b>	<b>145</b>
<b>APPENDIX IV. PERCEPTION TEST WORKSHEET .....</b>	<b>148</b>
<b>APPENDIX V. PERCEPTION TEST STIMULI .....</b>	<b>149</b>
<b>APPENDIX VI. PERCEPTION TEST STIMULI LINKING .....</b>	<b>151</b>
<b>APPENDIX VII. PRODUCTION TEST STIMULI .....</b>	<b>152</b>
<b>APPENDIX VIII. LINKING IN PRODUCTION TEST STIMULI .....</b>	<b>155</b>
<b>APPENDIX IX. PERCEPTION ITEM ANALYSIS .....</b>	<b>156</b>
<b>APPENDIX X. PRODUCTION ITEM ANALYSIS.....</b>	<b>158</b>
<b>APPENDIX XI. POST-TRAINING QUESTIONNAIRES .....</b>	<b>160</b>

## LIST OF FIGURES

Figure 1. Calvin and Hobbes comic with C-V linking humor .....	2
Figure 2. Classification of connected speech processes .....	10
Figure 3. NS pitch contour (top), learner's production (bottom).....	32
Figure 4. Annotated waveform display of <i>news items</i> .....	34
Figure 5. Visual feedback of <i>right away</i> .....	35
Figure 6. Visual feedback of <i>say it</i> .....	36
Figure 7. Waveform of <i>he's a great cook</i> showing linking between <i>he's</i> and <i>a</i> .....	39
Figure 8. Outline of the study procedure .....	49
Figure 9. Mark the link activity – AO group .....	54
Figure 10. Mark the link activity – AV group .....	55
Figure 11. Gapped dictation activity (same for both AO and AV groups).....	56
Figure 12. Mean perception scores for CG, AO and AV groups at pretest and posttest.....	62
Figure 13. Perception scores for pretest, posttest, and delayed posttest for AO and AV groups.....	64
Figure 14. Perception scores for posttest old and novel items for AO an AV groups.....	67
Figure 15. Mean production scores for CG, AO, and AV groups for pretest and posttest .....	84
Figure 16. Production scores for pretest, posttest, and delayed posttest for AO and AV groups.....	86
Figure 17. Production scores for posttest old and new parts for AO an AV groups. ....	90
Figure 18. Mean pretest and posttest scores for HF and LF words .....	91

## LIST OF TABLES

Table 1. Overview of classroom perception training studies of connected speech in ESL/EFL contexts .....	21
Table 2. Participant Overview .....	47
Table 3. Training unit activities .....	52
Table 4. Perception research questions, variables and analysis methods .....	60
Table 5. Pretest and posttest mean scores of linking perception for the all groups .....	62
Table 6. Results of two-way ANOVA test to examine long-term effect of treatment .....	63
Table 7. AO and AV one-way repeated measures ANOVA results .....	65
Table 8. Results of paired <i>t</i> -tests on perception posttests for old and novel contexts .....	67
Table 9. Item analysis of perception mean percentage scores .....	71
Table 10. Potential linking and phrasing in production stimuli.....	77
Table 11. Production test stimuli readability .....	78
Table 12. Production research questions, variables and analysis methods .....	82
Table 13. Pretest and posttest scores of linking production for all groups .....	84
Table 14. Results of two-way ANOVA test to examine long-term effect of treatment ...	86
Table 15. AO and AV one-way repeated measures ANOVA results .....	87
Table 16. Results of paired <i>t</i> -tests on production posttests for old and novel contexts.....	89
Table 17. Results of paired <i>t</i> test for HF and LF word linking .....	91
Table 18. Item analysis of selected production mean scores .....	97



Table 19. Post-training questionnaire questions for AO and AV groups .....	104
Table 20. Means and standard deviations of AO ad AV group perceptions of linking .....	106
Table 21. Summary of quantitative findings.....	114

## NOMENCLATURE

AO	Audio-only (group)
AV	Audio-visual (group)
CG	Control group
CSP	Connected speech process
C-V	Consonant-to-vowel (linking)
ESL	English as a Second Language
EVF	Electronic visual feedback
L1	First language
HF	High frequency
LF	Low frequency
L2	Second language
V-V	Vowel-to-vowel (linking)

## ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to those who helped me with the various aspects of conducting research and writing this dissertation. First and foremost, I thank my advisor, Dr. John Levis, for his guidance, patience, enthusiasm, and support throughout this process. John, I appreciate your ongoing dedication and the time you devoted to listening attentively, reading and re-reading drafts of this dissertation. Your insights and observations have inspired and helped me see things from new and various perspectives.

I also wish to express my appreciation to my committee members for their feedback during the course of this work. I further offer my appreciation to Dr. Janet Anderson-Hsieh for agreeing to be on my committee after her retirement. Her advice, thoughtful comments and lifetime experience in connected speech have been invaluable. I am sincerely grateful to the students of Iowa State University who volunteered to participate in the study, and indebted to all my friends and colleagues in the English Department at Iowa State University who helped me during various stages of this research. In addition, I thank YooRee Chun for her help in the statistical procedures.

On a more personal level, I thank my parents, Haifaa and Marwan and dedicate this dissertation to them. Despite the extremely difficult time and circumstances that my family have been living in for the last three years in oppressed Syria, they never ceased to support and encourage me and give me a reason to go on. I want to heartily thank my beloved husband, Kamal, and sweet daughter, Naya, for believing in me, cheering me up when I was overwhelmed, and providing a loving home. I am also grateful to my friend, Jane Scanlon, for her positive spirit and encouraging phone calls in the times I needed support the most. Without the love, patience, and help of these wonderful people, this project would not have been possible to complete.

## ABSTRACT

Despite the prevalence of connected speech processes in spoken English and their importance for communication, there has been relatively little research on them in the field of second language learning. This study investigated the effectiveness of using two methods for teaching linking to NNSs of English. Audio-visual (AV) feedback, integrating both audio and electronic visual feedback, and audio-only (AO) feedback were employed in the development of online materials to help improve learners' perception and production of consonant-to-vowel linking such as *givein* and vowel-to-vowel linking such as *seeit*. The long-term effectiveness of the instructional materials was also examined to see whether learners were able to retain improvement beyond the training period. In addition, the study explored whether improvement can transfer to novel contexts. The influence of high frequency and low frequency words on linking production was also investigated. Finally, students' perceptions of the use of AV and AO feedback were reported to guide future implementation of the materials.

Forty-five learners of English participated in the study with 15 students in each of the two experimental groups and the control group. A pretest, posttest and delayed posttest design was implemented to answer the research questions of the study. Students' dictation of old and new sentences and their audio recordings of two read texts were analyzed. Qualitative data consisted of students' answers to a post-training questionnaire.

Findings indicate that both types of training were effective in improving learners' perception and production of linking immediately and one month after training. Learners were also able to transfer the gained improvement to novel contexts. In addition, it was found that students' improvement of linking high frequency words was significantly greater than that of low frequency words, which were more difficult to link. The use of audio-visual training with waveforms resulted in better improvement in linking perception and production as well as more positive learner feedback than an audio-only training. The findings of the study hold a number of pedagogical implications for language teachers, material developers, and those interested in researching connected speech and the use of technology for language learning.

## CHAPTER I: INTRODUCTION

From a form-focused audio-lingual method to a fluency-focused communicative approach, pronunciation instruction has witnessed radical changes over the last 70 years. Some approaches have placed pronunciation at the forefront of instruction while others have completely ignored it or have assigned it a supplementary role in the classroom. By the 1990s, there was a gradual return to a more balanced approach that valued both accuracy and fluency. Suprasegmental features of spoken language (involving word stress, sentence stress, rhythm, and connected speech processes) were recognized as playing a critical role in the second language classroom (Gilbert, 2001; Pennington & Richards, 1986). They provide the framework for utterances and direct the listeners' attention to new and important information (Anderson-Hsieh, Johnson, & Koehler, 1992). In addition, prosody was claimed to have the greatest and fastest impact on the comprehensibility of learners' second language (L2) production (McNerney & Mendelsohn, 1992).

While many English language teacher training programs incorporate a pronunciation component, segmental aspects of phonology are often given more attention than suprasegmental aspects. This, in part, is because segmental phonology is easier to define, to identify and therefore to teach (Coniam, 2002). Tench (1996) suggested that instructors are much less confident in discussing suprasegmental features because they are features of "language in use rather than of language in units (like words)" (p. 2). However, since suprasegmentals may carry more of the overall meaning load than do segmentals, misunderstanding caused by suprasegmentals is apt to be a more serious issue than that caused by segmentals. Learners who use incorrect rhythm patterns or who do not connect words together are at best frustrating to the native listener. More seriously, if these learners use improper intonation contours, they can be perceived as abrupt, or even rude; and if the stress and rhythm patterns are too unfamiliar to listeners, the speakers who produce them may not be understood at all.

Connected speech processes (CSPs) are a significant suprasegmental aspect of pronunciation. They refer to phonological processes that take place in continuous chains of spoken language, such as linking, weak forms, elision, assimilation, and contraction. Because of

these processes, words spoken in isolation are quite different from those spoken in context. CSPs may change, delete, or add sounds to words (e.g., *aren't* instead of *are not*), leave their sounds relatively intact, or result in drastic changes to the connected words (e.g. *wouldjou* instead of *would you*). Linking, which is the primary focus of the present study, is a CSP that has the least effect on the pronunciation of the two connected words. It is what happens at word boundaries where two sounds combine while keeping their phonetic qualities. For instance, when the words *give in* are connected as *givin*, the phonetic qualities of the boundary sounds /v/ and /ɪ/ are kept unchanged.

The primary function of CSPs is to promote the regularity of English rhythm by compressing syllables between stressed elements and facilitating their articulation (Clark & Yallop, 1995). Although CSPs are found in all language registers, the more informal the speech register is, the more the citation forms of words are likely to change as a result of the interaction of these processes. Consequently, the pronunciation of connected speech may become a significant challenge to intelligibility both for native speakers (NS) of English and nonnative speakers (NNS). This, in turn, may impede communication by creating misunderstandings that are not always funny like the comic in Figure 1.

**Calvin and Hobbes** by Bill Watterson



Figure 1. Calvin and Hobbes comic with C-V linking humor (Watterson, 2008)

## **Statement of the Problem**

Connected speech processes are important for work and research in a number of areas, including teaching English as a Second Language (ESL), speech recognition software, and text-to-speech systems. Despite the prevalence of CSPs in spoken English and their importance for communication, there has been relatively little research on them in the field of second language learning. This is mainly because research in second language listening instruction has tended to focus more on the development of top-down skills than bottom-up skills like CSPs (Vandergrift, 2004). In addition, the complexity of experimental studies of casual speech production can render them intimidating to researchers (Shockey, 2003).

In the past few decades, there have been sporadic attempts to create more research and develop more interest in CSPs. In the field of L2 listening, Field (2003) and Rost (2006) called for greater attention to perceptual bottom-up skills and discussed methods for teachers and researchers to improve students' abilities to comprehend reductions. Correspondingly, many researchers have attempted not only to describe CSPs, but also to explore their prevalence and effect in native and nonnative speech (for example, Alameen, 2007b; Anderson-Hsieh, Riney, & Koehler, 1994; Barry, 1991; Brown & Kondo-Brown, 2006a; Hieke, 1984, 1987, 1989; Norris, 1994; Shockey, 2003; Temperley, 1987). In addition, several studies, especially recently, have examined the effectiveness of teaching CSPs on phonological awareness and listening comprehension (Brown & Hilferty, 1986a, 1986b; Carreira, 2008; Crawford, 2006; Fan, 2003; Henrichsen, 1984; Ito, 2006; Lee & Kuo, 2010; Liu, 2010; Matsuzawa, 2006; Wang, 2005; Yang, Lin, & Chung, 2009).

On the other hand, very few researchers have investigated the efficacy of CSP teaching materials on the speech production of NNSs of English (Kuo, 2009; Melenca, 2001; Sardegna, 2011). Most CSP studies have focused on features that are localized to certain words or phrases (e.g. assimilation and contractions), while only a few have investigated more pervasive features such as linking (Fan, 2003; Kuo, 2009; Melenca, 2001; Sardegna, 2011). This shows a need for more research on how linking instruction influences learners' speech production and perception. It is notable that despite the shortage of research on the instruction of CSPs, many pronunciation textbooks include a component that deals with such features. Moreover, although there have

been suggestions for including instructional materials for improving NNS perception and production of CSPs, their methods and materials remain largely untested.

In addition to the traditional methods of instruction, technology-enhanced materials utilizing electronic visual feedback (EVF) have been used to in teaching CSPs. Such materials have usually provided instantaneous feedback on the learner's speech production. The feedback is sometimes paired with the visual displays of NS model to guide learners to correct their own speech (Anderson-Hsieh, 1992; Chun, 1989; de Bot, 1983; Lane et al., 1988; Molholt, 1988). More research is needed to investigate the efficacy of using EVF and other technological approaches to teaching L2 speech perception and production.

In sum, more work is needed to develop language teaching materials to help NNSs understand and produce connected speech. Because many ESL teachers lack appropriate training in teaching pronunciation and because time in the classroom is usually limited and does not allow for much pronunciation practice, approaches that promote learner autonomy are especially needed. EVF approaches to teaching connected speech seem to be promising in that regard since they allow students to practice independently outside class. More research should be conducted to investigate and validate the efficacy of such materials. The present study addresses these gaps by examining two different approaches to teaching linking and evaluating their effectiveness in improving NNSs perception and production of connected speech.

### **Purpose of the Study**

The purpose of this study is to investigate the effectiveness of using two methods for teaching linking to NNSs of English. Audio-visual (AV) feedback, integrating both audio and electronic visual feedback, and audio-only (AO) feedback were employed in the development of online materials to help improve learners' perception and production of consonant-to-vowel (C-V) linking such as *givein* and vowel-to-vowel (V-V) linking such as *seeit*. The long-term effectiveness of the instructional materials was also examined to see whether learners were able to retain improvement beyond the training period. In addition, the study explored whether improvement can transfer to novel contexts; in other words, it probed learners' ability to link and understand linked words in new and unpracticed utterances. The influence of high frequency and



low frequency words on linking production was also investigated. Finally, students' perceptions of the use of AV and AO feedback were reported to guide future implementation of the materials.

Participants for the research study were international students enrolled in Engl 99L: Strategies for Listening at Iowa State University. The majority of the students were undergraduates with a small percentage of graduate students. In addition, the participants included a large number of Chinese native speaking students due to the nature of the international student population at Iowa State University. The study took place in three intact classes where every section was randomly assigned a type of treatment: the control group (CG) received no treatment, the AO group received training with audio-only feedback, and the audio-video (AV) group received training with electronic audio-visual feedback.

A quasi-experimental design with pretest, posttest, delayed posttests and a control group was utilized to investigate the immediate and long-term effects of AO and AV training on learners' perception and production of linking. In addition, the study examined the efficacy of training on novel contexts, or in other words, the extent of generalization and transfer of improvement to new and unpracticed utterances. The impact of word frequency on linking production was also explored. The quantitative phase of the study was followed by a qualitative component to add depth to the quantitative results. Survey questions elicited responses from participants regarding their experiences using AV or AO training materials and their suggestions for improving such approaches to teaching linking.

## **Research Questions**

In sum, the study attempts to answer the following research questions about the instruction of linking perception and production:

1. How effective are AO and AV training in improving linking perception and production for non-native learners of English?
2. How is the improvement caused by AO and AV training sustained over time?

3. How do AO and AV training generalize to novel perception and production contexts?
4. What is the impact of word frequency on linking production? Are low frequency words more difficult to link to and from than high frequency words? Do learners improve better after treatment in HF contexts than in LF contexts?
5. What are learners' perceptions of the use of audio-only and electronic visual feedback in teaching English linking?

### **Significance of the Study**

The state of affairs described above suggests that there are some gaps in the field of L2 pronunciation instruction, specifically in the teaching of connected speech. In an effort to help L2 learners, teachers, researchers and, material designers, this study aimed to test the immediate and long-term effectiveness of two different approaches to teaching linking perception and production. Additionally, it investigated the influence of context novelty and word frequency on linking.

Findings of the study may help improve the development of CSP instructional materials. The hypothesis that AV training involving waveform displays improves linking perception and production, if supported, would contribute to the quest for more appropriate, efficient, and innovative pedagogical tools. Such tools can help learners improve their pronunciation and encourage teachers to incorporate more connected speech instruction in their pronunciation teaching repertoire. In addition to evaluating two CSP instructional approaches, the findings of this study will help educators set teaching priorities in the classroom. By knowing which linking type or word categories are more efficient to teach, teachers and material designers can plan to incorporate those items in the syllabus and avoid overwhelming learners with rules and contexts they are less likely to encounter. Finally, this study may encourage researchers to expand on CSP research by investigating other CSP aspects and advanced pedagogical tools.

### **Organization of the Dissertation**

The dissertation is divided into seven chapters. In chapter one I have presented the research justification, the purpose of the study and its significance. In chapter two, I review

relevant literature in the areas of CSP effects on listening perception and production, the effectiveness of instruction on listening perception and production, the effect of word frequency on connected speech, and the use of EVF in the second language classroom. I discuss the theoretical underpinning of the research, previous work in the area, and unique methodologies. Chapter three describes the methods, including participants in the study, training units for experimental groups, and study procedures. Due to the dual goal of the study, the rest is organized into two major perception and production chapters. Each chapter includes a description of the corresponding test stimuli, analysis procedure, results, and discussion. Chapter six contains a description of the post-training questionnaire tool that was used to survey learners' opinions about the training. The chapter also presents the qualitative findings of the questionnaire and discusses them in light of the quantitative results. Finally, chapter seven concludes the dissertation by discussing the pedagogical implications of the findings and study limitations, and makes suggestions and recommendations in light of the results.

## **CHAPTER II: LITERATURE REVIEW**

This chapter contextualizes the present study within relevant research literature on connected speech. It starts by an overview of CSP definitions, features, and social and linguistics factors that can influence them in English. It then discusses linking features and types while providing examples of the phenomena. Literature about teaching connected speech perception and production is then reviewed with special focus on studies on linking. Additionally, the section looks at how English as Second Language (ESL) materials deal with the topic and discusses how EVF has been used to teach suprasegmental features.

### **Connected Speech Processes**

Hieke (1987) defined connected speech processes as “the changes which conventional word forms undergo due to the temporal and articulatory constraints upon spontaneous, casual speech” (p. 41). In other words, they are the processes that words undergo when their border sounds are blended with neighboring sounds. They include reduction, elision, assimilation, intrusion, linking, liaison, and contraction, among others. This means that pronunciation of consonant and vowel sounds in running speech often differs from the pronunciation of the sounds when words are uttered in isolation, that is, in citation form (Lass, 1984). Citation form pronunciations occur in isolated words under heavy stress or in sentences delivered in a slow, careful style. By contrast, connected speech forms often undergo a variety of modifications which cannot always be predicted by applying phonological rules (Anderson-Hsieh et al., 1994; Hieke, 1987; Lass, 1984; Temperley, 1987). It may be that all languages have some form of connected speech processes, as Pinker (1995) claims:

In speech sound waves, one word runs into the next seamlessly; there are no little silences between spoken words the way there are white spaces between written words. We simply hallucinate word boundaries when we reach the edge of a stretch of sound that matches some entry in our mental dictionary. This becomes apparent when we listen to speech in a foreign language: it is impossible to tell where one word ends and the next begins (pp. 159-160).

The primary function of CSPs in English is to promote the regularity of English rhythm by compressing syllables between stressed elements and facilitating their articulation so that regular running speech timing can be maintained (Clark & Yallop, 1995). For example, certain closed class words such as prepositions, pronouns, and conjunctions are rarely stressed, and often appear in a weak form in these unstressed contexts. Consequently, they are ‘reduced’ in a variety of processes to preserve the rhythm of the language. Reducing speech can also be attributed to the law of economy where speakers economize on effort, avoiding, for example, difficult consonant sequences by eliding sounds (Field, 2003a). The organs of speech, instead of taking a new position for every sound, tend to connect sounds together with the purpose of saving time and energy (Clarey & Dixon, 1963).

### **Connected speech features**

In discussing connected speech, two issues cannot be overlooked: differences in terminology and the infrequency of relevant research. Not only do different researchers and material designers use different terms for CSPs (e.g., sandhi variations, reduced forms, absorption), they also do not always agree on how to classify them. In most textbooks and research, CSPs are presented as a list of independent processes, where every one is enlisted with a definition and different examples. However, this can be an oversimplified picture of what really happens to words when spoken together, especially in spontaneous speech. In real-time communication, words may be radically modified in ways not conventionally dealt with in textbooks (Cauldwell, 2013). To account for such processes (or combination of processes), Alameen and Levis (in press) classified CSPs into six main categories: linking, deletion, insertion, modification, reduction and multiple processes (Figure 2). Every category includes processes that share similar characteristics but are applied differently.

This classification model recognizes the complexity of connected speech, a contributing factor in the paucity of CSP research. Conducting experimental studies of connected speech can be intimidating to researchers because “variables are normally not controllable and one can never predict the number of tokens of a particular process one is going to elicit, which in turn makes the application of statistical measures difficult or impossible” (Shockey, 2003, p. 109). As a

result, only a few people have worked on the issue and those few have done so only sporadically (Brown & Kondo-Brown, 2006a).

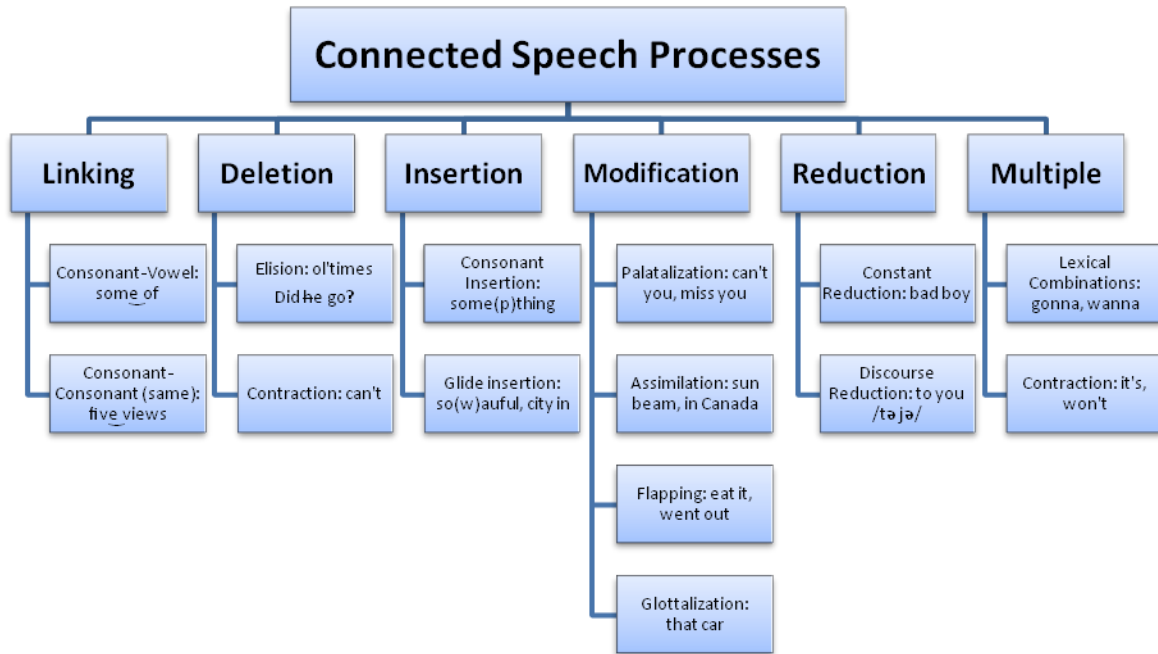


Figure 2. Classification of connected speech processes (Alameen & Levis, in press)

Despite such complexity, several studies have investigated an array of connected speech phenomena in native speaker production, and attempted to quantify their characteristics. These studies examined processes such as assimilation and palatalization (Barry, 1991; Shi, Gick, Kanwischer, & Wilson, 2005), deletion (Norris, 1994), contraction (Scheibman, 2000), liaison (Allerton, 2000), linking (Alameen, 2007b; Hieke, 1987; Temperley, 1987; Trammell, 1999) and nasalization (Cohn, 1993). Such studies provide indispensable background for any research in L2 perception and pronunciation.

It appears that certain social and linguistic factors affect the frequency, quality, and contexts of CSPs. Lass (1984) attributes CSPs to the immediate phonemic environment, speech rate, the formality of the speech situation and other social factors, such as social distance. Hieke (1984) distinguished two styles of speech: casual everyday style and careful speech used for certain formal occasions, such as presentations. According to Hieke, in casual spontaneous speech, speakers pay less attention to fully articulating their words, hence reducing the

distinctive features of sounds while connecting them. Similarly, when examining linking for NS and NNS of English, Anderson-Hsieh et al. (1994) found that style shifting influenced the manner in which speakers link their words. In their study, NSs and NNSs performed more linking in spontaneous speech tasks than those involving more formal sentence reading.

However, other researchers found that while there was some evidence that read speech was less reduced, unscripted and scripted speech show great phonological similarity (Alameen, 2007b; Shockey, 1974). The same processes apply to both styles and very nearly to the same degree. Native speakers do not seem to know that they are producing speech which differs from citation form. In Alameen (2007b), NNSs as well as NSs of English did not have significant differences between their linking performance in text reading and spontaneous speech tasks, which indicates that a change in speech style may not entail a change in linking frequency. Furthermore, Shockey (2003) noted that many CSPs occur in fast speech as well as in slow speech, so “if you say ‘eggs and bacon’ slowly, you will probably still pronounce ‘and’ as [m], because it is conventional - that is, your output is being determined by habit rather than by speed or inertia” (p. 13).

Since read scripted speech has been shown to be similar to unscripted spontaneous speech in terms of linking frequency, scripted speech is the style to be used in the present study for data collection and training purposes. An added advantage is that it is much easier to control for variables such as linking frequency, difficulty level, and pausing in scripted speech. One problem that could arise from using scripted speech, though, is some readers’ tendency to read words in citation form, not connected forms. This, however, seems to influence processes that cause a change in the nature of sounds, such as deletion and reduction, but to a lesser degree in linking, where boundary sounds keep their characteristics. Additionally, read speech may better approximate the conditions of spontaneous speech by native speakers because the learners are relieved of the burden of worrying about vocabulary choice, correct grammar, and sentence formation (Melenca, 2001). Consequently, they may speak more fluently with a script than in unscripted conversation (Alameen, 2007b). In this study, the participants will read all stimuli twice to eliminate hesitation due to text novelty.

Other factors, such as social distance, play a role in determining the frequency with which such processes happen (Anderson-Hsieh et al., 1994). When the speaker and the listener both belong to the same social group and share similar speech conventions, the comprehension load on the listeners will be reduced, allowing them to pay less attention to distinctive articulation. Indeed, in formal situations some CSPs are completely acceptable (e.g., linking), while other forms are less acceptable (e.g. *gonna* for *going to*).

Variation in degree is another feature that characterizes CSPs. Many researchers think of connected speech processes in clear-cut definitions; however, speakers do not always produce a specific CSP in the same way. A large study of CSPs was done in the University of Cambridge, results of which appeared in a series of articles (Barry, 1984, 1985, 1991; Wright, 1986). The results showed that most CSPs produce a continuum rather than a binary output. For instance, if the process of contraction suggests that *do not* should be reduced to *don't*, we often find, phonetically, cases of both expected variations and a rainbow of intermediate stages, some of which cannot be easily detected by ear. Such findings are insightful for CSP instruction since they help researchers and teachers decide on what CSP to give priority depending on the purpose and speech style. They also provide a better understanding of CSPs that may facilitate the development of CSP instructional materials.

## **Linking**

### **Definition of linking**

The term *linking*, also known as attraction, juncture, and transition, has been used in different ways. It can refer to adjustments speakers make between words in connected speech in general (Goodwin, 2001), in other words, CSPs in general. However, the more specific agreed-upon use of the term (Brown & Kondo-Brown, 2006b; Celce-Murcia, Brinton, Goodwin, & Griner, 2010; Gimson, 1989; Hieke, 1984) refers to what happens to sounds at word boundaries in one thought group when two words are joined by connecting the last sound of one word to the first sound of the following word. The two connected sounds can (a) keep their phonetic qualities as in consonant-to-vowel (C-V) linking (e.g. *face it*), (b) have an extra glide inserted in between them as in vowel-to-vowel (V-V) linking (e.g. *blue ink*), or (c) be combined in one longer sound



as in same consonant-to consonant (C-C) linking (e.g. *can name*), or (d) have changes in segment identity, as in different consonant-to-consonant linking where the first consonant may not be released or aspirated (e.g. *let down*).

It appears that linking occurs due to the avoidance of hiatus, in other words, the avoidance of leaving sounds at word boundaries without a transition between them (Allerton, 2000; Anderson-Hsieh et al., 1994; Hieke, 1984). Since initial vowels are preceded by glottal stops, such a “pitch-interruptive feature becomes a primary motivation for linking” (Hieke, 1984, p. 346). The glottal stop /ʔ/ is a plosive created by complete closure and then opening of the glottis (vocal folds). Although it is a consonant sound in other languages, in English it generally appears as an allophone of /t/ and /d/, as in ‘*cotton*’. In order to avoid an interruption and arrive at a non-initial vowel structure without a glottal stop, syllabic restructuring (or resyllabification) takes place in casual speech. For example, *find out* is pronounced more like *fine doubt*. However, initial vowels may sometimes be kept and produced with a slight glottal onset under certain conditions, such as special stress assignment and after pauses (Hieke, 1984).

In the light of this discussion, glottal onset on initial vowels is used as a marker for unlinked words. In other words, hearing a glottal stop between two words indicates that these two words are not linked. This description is very helpful in establishing a criterion to identify linking, especially when it is difficult for raters to detect a break of linkage by ear. In this study, a linking pair (the two linked words) is considered to be unlinked when a glottal stop is detected between the two words. As a result, this study adopts the following definition of linking: *Linking is connecting the boundaries of two words while keeping their phonetic qualities without detecting a glottal stop between the two words.*

### **Types of linking**

According to the sounds that meet at word boundaries (i.e. consonants and vowels), linking can be categorized into three main types.

### 1. Consonant-to-vowel (C-V) Linking

Consonant-to-vowel linking (C-V) takes place when the final consonant of a word is followed by a vowel at the beginning of the next word in the same thought group. The final consonant is often pronounced as a medial consonant, a consonant occurring in the middle of a word, such as [s] in 'face it' (Hieke, 1984). Therefore, when a word terminating in a consonant cluster (CC + V sequence) or a single consonant (VC+V sequence) is followed by a word or syllable commencing with a vowel, the final consonant of the cluster is often pronounced as part of the following syllable (Celce-Murcia, Brinton, & Goodwin, 1996). For instance, in CC+V sequence *found out* is pronounced as [faʊn daʊt] rather than [faʊnd aʊt], and in VC+V sequences *give in* is pronounced as [gi \_vɪn] rather than [gɪv ɪn] (where the symbol \_ represents a link between words, in other words, no auditory space between them).

What complicates the situation for the listener is that, after resyllabification, words sometimes acquire false boundary cues (Field, 2003a). Thus, in *left in*, the /t/ might be lightly aspirated, suggesting that it is word-initial. Voiceless stops, such as /t/, are aspirated when they are syllable initial. Furthermore, resyllabification makes it difficult for L2 learners to find the word boundary. When they hear the utterance [meɪdaʊt] for example, it would be difficult to decide whether it means *may doubt* or *made out* without enough context. The present study will treat all resyllabification and ambisyllabic segments (i.e, a consonant that cannot be assigned exclusively to one syllable or another but is shared by both) at word boundaries as incidents of C-V linking since, regardless to which syllable the ambisyllabic consonant is assigned, the link still takes place between the two border sounds.

A special note needs to be made about /h/ deletion in C-V linking. /h/-deletion happens when the final consonant of a word is connected to the initial /h/ of the following word. The /h/ sound tends to disappear and the final consonant is linked with the vowel following /h/. This usually takes place when the word starting with /h/ is a pronoun (e.g., him, her, his, he) or auxiliary verb (e.g., have, has). For instance, *tell him* is pronounced as [tɛlɪm].

## 2. Vowel-to-vowel (V-V) Linking

Vowel-to-vowel (V-V) linking occurs when a word that ends in a high and mid-tense vowel is followed by a word that begins with a vowel. Speakers, thereafter, insert a junctural glide, a very short /<sup>w</sup>/ or /<sup>j</sup>/ sound, to link the two vowels together and avoid a gap between the sounds (Celce-Murcia et al., 2010; Cruttenden, 2008). The choice of the junctural glide depends on the vowel at the end of the first word. If the first word ends with a high or mid front vowel, then the linking sound will be /<sup>j</sup>/ as in *my ear* [maɪ<sup>j</sup>ɪr]. However, If the word ends with a high back or mid vowel, then the two words are linked with /<sup>w</sup>/ as in *now I* [naʊ<sup>w</sup>aɪ]. Other vowels are usually smoothly linked without a linking sound (Hewings & Goldstein, 1998). It is worth mentioning that the junctural [j] glide is different from the phoneme [j] and [w] is different from [w] in that the finishing point of the diphthong is not sufficiently prominent and the glide is not long enough to be identical to the full sound. This can be noted in the opposition between *my ear* [maɪ<sup>j</sup>ɪr] and *my year* [maɪ<sup>j</sup>ɪr] (Cruttenden, 2008).

## 3. Consonant-to-consonant (C-C) Linking

Consonant-to-consonant (C-C) linking can take place when two identical consonants meet at word boundaries and consequently are pronounced as one slightly prolonged sound, as in *can name* [kəneɪm]. When the two border consonants are otherwise different, the final consonant of a word is then released at the beginning of the following word, such as the unaspirated, unreleased /t/ in *what she* (Anderson-Hsieh, 1994). The latter C-C condition only applies when the first consonant is a stop and may include more variety and change than other types of linking. For instance, the /t/ in *great clips* [greɪ<sup>ʔ</sup>klɪps] can be glottalized at times. Because of its complexity and variety, C-C linking will not be discussed in the study and the training materials. More time should be dedicated to training students on C-C linking (Melenca, 2001), but it is beyond the scope of my study.

## **Connected Speech Perception**

In spoken language situations, frustrating misunderstandings in communication may arise because NSs do not pronounce English the way L2 learners are taught in the classroom. L2

learners' inability to decipher spontaneous speech comes from the fact that they develop their listening skills based on the adapted English speaking styles they experience in an EFL class. In addition, they are often unaware of the differences between citation forms (i.e., clear pronunciation of a word when it is stressed or pronounced in isolation, out of context) and modifications in connected speech (Shockey, 2003). When listening to authentic L2 materials, Brown (1990) claims an L2 learner

Will hear an overall sound envelope with moments of greater and lesser prominence and will have to learn to make intelligent guesses, from all the clues available to him, about what the probable content of the message was and to revise this interpretation if necessary as one sentence follows another – in short, he has to learn to listen like a native speaker (p. 4).

A part of the L2 listener's problem can be attributed to the fact that listening instruction has tended to emphasize the development of top-down listening processes over bottom-up processes (Field, 2003a; Vandergrift, 2004). However, in the past decade, researchers have increasingly recognized the importance of bottom-up skills, including CSPs, for successful listening (Rost, 2006). In the first and only book dedicated to researching CSPs, Brown and Kondo-Brown (2006b) noted that despite the importance of CSPs for learners, little research on their instruction has been done, and stated that the goal of their book is to “kick-start interest in systematically teaching and researching connected speech” (p. 6). This, to an extent, seems to have started to take an effect with more CSPs studies and theses conducted in the last few years especially in Taiwan (e.g., Kuo, 2009; Lee, 2012; Wang, 2005) and Japan (e.g., Crawford, 2006; Matsuzawa, 2006). The next section will discuss strategies NSs and NNSs use to understand connected speech, highlight the effect of CSPs on L2 listening and review the literature on the effectiveness of CSPs perceptual training on listening perception and comprehension. It is to be noted that what is referred to as ‘linking perception’ or ‘CSP perception’ is the learners’ metalinguistic awareness of linking or the CSP(s).

### **Speech segmentation**

A good place to start addressing L2 learners' CSPs problems is by asking how native listeners manage to locate word boundaries and successfully segment speech. Some models of

speech perception propose that specific acoustic markers are used to segment the stream of speech (e. g., Nakatani & Dukes, 1977). In other models, listeners are able to segment connected speech through the identification of lexical items (McClelland & Elman, 1986). Cutler (1990) suggests that native listeners use a strong-syllable strategy, based on the premise that each stressed syllable marks the beginning of a new word. Furthermore, Rost (2006) describes two complementary bottom-up phonological processes that assist the listener: feature detection and metrical segmentation. Feature detection enables the listener to decode speech into linguistic units and is influenced by his/her L1. Metrical segmentation refers to the use of stress and timing rules to segment incoming speech into words. In connected speech, the listener compares a representation of the actual speech stream to stored representations of words. Here, the presence of CSPs may create lexical ambiguity due to the mismatch between the lexical segments and their modified phonetic properties. For experienced listeners, however, predictable variation does not cause a breakdown in perception (Gaskell, Hare, & Marslen-Wilson, 1995).

Rather than using phonological context to decipher connected speech, NNSs depend heavily on syntactic-semantic information taking in a relatively large amount of spoken language to process and thereby introducing a processing lag instead of processing language as it comes in (Shockey, 2003), much as predicted in Brown (1990). L2 learners' speech segmentation is primarily led by lexical cues pertaining to the relative usage frequency of the target words, and secondarily from phonotactic cues pertaining to the alignment of syllable and word boundaries inside the carrier strings (Sinor, 2006).

### **Effects of CSPs on listening perception and comprehension**

The influence of connected speech on listening perception has been investigated in several studies (Brown & Hilferty, 1986b; Henrichsen, 1984; Ito, 2006). These studies also show how reduced forms in connected speech can interfere with listening comprehension. Evidence that phoneme and word recognition are indeed a major source of difficulty for low-level L2 listeners comes from a study by Goh (2000). Out of ten problems reported by second-language listeners in interviews, five were concerned with perceptual processing. Low-level learners were found to have markedly more difficulties of this kind than more advanced ones.

In a pioneer study in CSP research, Henrichsen (1984) examined the effect of presence and absence of CSPs on ESL learners' listening comprehension skills. He administered two dictation tests to NNS of low and high proficiency levels and NSs. The results confirmed his hypothesis that reduced forms in listening input would decrease the saliency of the words and therefore make comprehension more difficult for ESL learners. Comprehending the input with reduced forms, compared to when the sentences were fully enunciated, was more difficult for both levels of students meaning, that connected speech was not easy to understand regardless of the level the students were in.

Ito (2006) further explored the issue by adding two more variables to Henrichsen's design: modification of sentence complexity in the dictation test and different types of CSPs. She distinguished between two types of reduced forms, lexical and phonological forms. Her assumption was that *lexical reduced forms* (e.g., *won't*) exhibit more saliency and thus would be more comprehensible compared to *phonological forms* (e.g., *she's*). As in Henrichsen's study, the nonnative participants scored significantly higher on the dictation test when reduced forms were absent than when they were present. Furthermore, NNSs scored significantly lower on the dictation test of phonological forms than that of lexical forms, which indicated that different types of reduced forms did distinctively affect comprehension. Considering the effects of CSPs on listening perception and comprehension and the fact that approximately 35% of all words can be reduced in normal speech (Bowen, 1975, cited in Cahill, 2006), perceptual training should not be considered a luxury in the English language classroom.

### **Effectiveness of CSP perceptual training on listening**

Since reduced forms in connected speech cause difficulties in listening perception (i.e., listening for accuracy) and comprehension (i.e., listening for content), several research studies have attempted to investigate the effectiveness of explicit instruction in connected speech on listening. This section will review eight classroom studies (Brown & Hilferty, 1986b; Carreira, 2008; Crawford, 2006; Kuo, 2009; Lee & Kuo, 2010; Matsuzawa, 2006; Ting & Kuo, 2012; Wang, 2005), describing their training materials and instrument and highlighting their findings. Table 1 summarizes the eight studies. It is to be noted that this is not a comprehensive list, but it includes the most prominent and/or accessible research studies in the field.

Since Henrichsen's revelation that features of CS reduced perceptual saliency and affected ESL listeners' perception, the above-mentioned researchers have responded to the need of exploring the effectiveness of teaching CS to a variety of participants. In addition to investigating whether L2 perceptual training can improve learners' perceptual accuracy of CSPs, some of the researchers wanted to examine the extent to which such training can result in improved overall listening comprehension (namely Brown & Hilferty, 1986b; Carreira, 2008; Lee & Kuo, 2010; Wang, 2005). The types of CSPs/reductions that could be taught effectively with perceptual training or that are more difficult for students were also considered in some studies (Crawford, 2006; Kuo, 2009; Ting & Kuo, 2012). Furthermore, students' attitudes toward listening difficulties, types of reduced forms, and reduced forms instruction were surveyed (Carreira, 2008; Kuo, 2009; Matsuzawa, 2006; Wang, 2005).

The range of connected speech processes explored in those studies was not comprehensive. Some focused on teaching specific high frequency reductions, which are word combinations undergoing various CSPs and appearing more often in casual speech than others; for instance 'gonna' for 'going to', palatalization in 'couldja' instead of 'could you', and C-V linking and contraction in 'if it' (Brown & Hilferty, 1986b; Carreira, 2008; Crawford, 2006; Matsuzawa, 2006). On the other hand, some were more restrictive and researched certain processes, such as C-V linking, palatalization, and assimilation (Kuo, 2009; Lee & Kuo, 2010; Ting & Kuo, 2012; Wang, 2005). These studies trained the participants to recognize the CSP general rule using a great number of reduction examples, instead of focusing on a limited number of examples and teaching them repeatedly.

The participants recruited in the previous studies included graduate and undergraduate students (Brown & Hilferty, 1986b; Carreira, 2008; Crawford, 2006; Ting & Kuo, 2012; Wang, 2005), school students (Kuo, 2009; Lee & Kuo, 2010) and even business people (Matsuzawa, 2006). What is common among these studies is that most participants were Asian EFL learners, who mainly had Japanese and Chinese as their L1. I could not find any CSP studies that dealt with other populations, which suggests that interference of such native languages may cause CSP problems and consequently more researcher interest.

The research design of the studies was similar since they all were conducted in classrooms with intact classes, except for Brown and Hilferty (1986b), where participants were randomly assigned to groups. Two studies did not have a control group (Carreira, 2008; Matsuzawa, 2006), while the rest had an experimental group trained on CS perception and a control group trained on other issues. The absence of a CG may affect the validity of these studies. Only Ting and Kuo (2012) compared the effectiveness of two instructional methods, thus using two experimental groups. A pretest-posttest design was implemented in all the studies to investigate post-instructional changes; however, none of them examined the long-term effects of CSP instruction on learners' perceptual accuracy.

Due to the lack of CSP teaching materials in general, most researchers developed their own instructional materials and borrowed examples from pronunciation textbooks. Only Carreira (2008) and Ting and Kuo (2012) used exclusively textbook materials such as *Hit Parade Listening* (Kumai & Timson, 2003) in their training. On average, training sessions lasted 2 to 4 hours, except for the longer treatment of Carreira (2008) and Kuo (2009) of 45 and 18 hours, respectively. Short and simple isolated sentences were mostly used for training with sporadic incorporation of short dialogs. However, three researchers used more authentic and contextualized materials taken from popular songs (Carreira, 2008; Ting & Kuo, 2012) and the movie *You've Got Mail* (Wang, 2005). All studies employed an explicit rule instruction approach that started with a presentation of the CSP with examples, followed by practice exercises, primarily dictation. The strength of dictation, as Crawford and Ueyama (2011) claimed, is that it simulates the cognitive processing that learners engage in when they encounter connected speech. Complete sentence dictation and cloze dictation were the most common types of perceptual exercises used. In cloze dictation, learners only write the two (or one) words affected by CSPs, for example 'Where ..... go?' where the connection of the two missing words, 'did' and 'you' involves the process of palatalization. Lee and Kuo (2010) were the only ones to venture into a different approach of training. They compared the effectiveness of teaching CSPs using explicit rule instruction and communicative tasks, such as information gap and problem solving.



Table 1. Overview of classroom perception training studies of connected speech in ESL/EFL contexts

Study	Goal	CSPs Treated	Research Design	Time for Training	Training Materials	Testing Instrument	Findings
<b>Brown &amp; Hilferty, 1986b</b>	LC & CSP	A number of reductions	EG (N=16), CG (N=16). ESL Chinese graduate students.  Random assignment.	4 weeks, daily 10-minute lessons (3 hours)	Self-developed sentences. Presentation of the forms, check comprehension with dictation exercises.	Sentence dictation test, LC test (ESLPE)	EG significantly increased their ability to identify and write down reductions by 32% over the pretest, but the improvement was not significant for the LC test.
<b>Wang, 2005</b>	LC & CSP	Elision, assimilation, contraction, linking	EG (N=37), CG (N=35). EFL Taiwanese freshman students.  Intact classes.	7 weeks (every 2 weeks), 30-minute lessons (3.5 hours)	Self-developed units based on video clips from a movie.  Comprehension check, dictation, CSP description, short dictation.	Sentence dictation test, LC test (GEPT)	EG significantly increased their awareness of reduced forms. EG did not outperform the CG in the listening comprehension posttest.
<b>Crawford, 2006</b>	CSP	A number of reductions	EG (N=23), CG (N=26). EFL Japanese freshman students.  Intact classes.	7 weeks, 15-minute lessons (2 hours)	Self-developed sentences.  Description, examples, cloze dictation.	Cloze dictation	For most of the reductions, gains were over 30% in the posttest.

LC stands for Listening comprehension; EG for experimental group.

Table 1 (Continued)

Study	Goal	CSPs Treated	Research Design	Time for Training	Training Materials	Testing Instrument	Findings
<b>Matsuzawa, 2006</b>	CSP	A number of reductions	EG (N=20), no CG. EFL Japanese business people.	4 weeks, 30-minute 7 lessons (3.5 hours)	Borrowed from Hagan, (2000). Explanation, sentence and cloze dictation	Sentence dictation test	Participants had difficulty in comprehending reduced forms. In the post-test, the participants made significant improvement. A positive correlation was found between the participants' English proficiency level and their comprehension of reduced forms.
<b>Carreira, 2008</b>	LC & CSP	A number of reductions	EG (N=19), no CG. EFL Asian students in Japan.	30 weeks, 90 minute-lesson (45 hours)	Songs chosen from <i>Hit Parade Listening</i> . Description, practice, cloze dictation	Sentence dictation test, TOEIC listening section for LC	Participants made significant improvement in recognizing spoken words. EG did not outperform the CG in the TOEIC listening comprehension posttest.
<b>Kuo, 2009</b>	CSP	C-V linking, /h/ deletion, C-C linking, V-V linking	EG (N=33), CG (N=32). EFL Taiwanese six graders.  Intact classes.	14 weeks, 40-minute lessons, twice a week (18 hours)	Self-developed worksheets. Explanation, modeling, repetition, practice, song teaching, production recording	Sentence reading, listen and circle, mark linking, cloze dictation	EG significantly improved their speech production and developed phonological awareness.

Table 1 (Continued)

Study	Goal	CSPs Treated	Research Design	Time for Training	Training Materials	Testing Instrument	Findings
<b>Lee &amp; Kuo, 2010</b>	LC & CSP	C-V linking, C-C reduction and palatalization	EG1 (explicit instruction), EG2 (communicative instruction), CG. N=30-32 each. EFL Taiwanese nine graders.  Intact classes.	3 weeks, daily 15-minute lessons, (4 hours)	Self-developed.  Explicit CS group: description, cloze dictation.  Communicative group: 15 activities with CSP in topic sentence.	Cloze dictation, LC test	The communicative group performed significantly better than the other groups on the cloze test. Explicit instruction group did not perform significantly better than the CG on the cloze test; no significant improvement was found on the listening comprehension among the three groups.
<b>Ting &amp; Kuo, 2012</b>	CSP	C-C linking (elision), C-V linking, /h/-deletion, contraction, palatalization, flapping	EG, CG (N not reported). EFL Taiwanese sophomore English majors.  Intact classes.	6 40-minute lessons (4 hours)	Same as Carreira, 2008	Sentence dictation test (read with presence and absence of CSPs)	EG made significant improvement in recognizing spoken words.

To assess improvement in perceptual accuracy, listening tests based on sentence dictation or cloze dictation were primarily used, while self-developed (Lee & Kuo, 2010) or adapted listening comprehension tests (Brown & Hilferty, 1986b; Carreira, 2008; Wang, 2005) were used to evaluate improvement on listening comprehension. With respect to their findings, studies on the effectiveness of CSP instruction showed mixed results. Most studies found that explicit CSP instruction was effective in improving learners' perception of CSPs and therefore should be taught in the classroom. It is only in Lee and Kuo (2010) that this approach failed to promote significant gains, where the communicative group outperformed the other two groups. Such a finding should encourage the use of more communicative materials when teaching CSPs. All studies failed to show significant improvement in listening comprehension. The inconsistent research findings require more empirical studies to clarify the effectiveness of connected speech instruction. Finally, research showed that the CSPs least amenable to perceptual training were /h/ deletion, as in 'e' for 'he' and flapping (where /t/ or /d/ is pronounced as an alveolar flap /r/), as in the /t/ in 'but I' (Crawford, 2006; Ting & Kuo, 2012). In addition, V-V linking, with a larger degree of variance than other types of linking, posed more problems to the experimental group (Kuo, 2009). These CSPs appear to be challenging for learners, and may require either additional training or a different type of training.

Results of the previous studies generally indicate that CSP instruction facilitated learners' perception of connected speech. However, the studies failed to address the long-term effects of such training on learners' perceptual accuracy. Moreover, no study has investigated generalization and transfer of improvement to novel contexts which indicates that improved abilities could extend beyond the training to natural language usage. This current study aims to address some of these shortcomings by examining the long-term effectiveness of CSP instruction and its spread to novel contexts. In addition to explicitly teaching CSPs through audio samples, this study introduced electronic visual feedback (EVF) as a method to raise learners' awareness of CSPs in spoken language. What is more, while most studies have used isolated sentences to teach CSPs, authentic materials taken from movies as well as ESL materials were incorporated in this study to present real life listening contexts (Herron & Seay, 1991). Finally, instead of surveying a number of CSPs, the learners in the study were presented with only C-V and V-V linking to focus their efforts on a smaller set of processes.

## **Connected Speech Production**

Temperley (1987) suggests that “closer examination of linking shows its more profound effect on English pronunciation than is usually recognized, and that its neglect leads to misrepresentation and unnatural expectations” (p. 65). However, the study of linking – and connected speech phenomena in general – has been marginalized within the field of speech production. This section discusses connected speech production in NS and NNS speech highlighting its significance and prevalence, and demonstrating the effectiveness of training in teaching CS production.

### **Connected speech in NS and NNS production**

Hieke (1984, 1987), Anderson-Hsieh et al. (1994) and Alameen (2007b) investigated aspects of connected speech production of American English, including linking, and compared them to those of nonnative speakers of English. In a series of small-scale studies, Hieke (1984, 1987) investigated the actual prevalence and distribution of selected CSPs in native and nonnative speech. Samples of spontaneous, casual speech were collected from NS (n=12) and NNS (n=29) participants according to the paraphrase mode, that is, they retold a story heard just once. C-V linking, alveolar flapping, and consonant cluster reduction were considered representative of major connected speech categories for the purposes of this pilot. A discrepancy was found between NNS and NS actualizations of about 30% for all three CSPs. The results showed that C-V linking in casual speech is present at the rate of 12 links per 100 syllables, which showed that linking can be considered as a rule and not a tendency in English. Hieke (1987) concluded that these phenomena could be considered “prominent markers of running speech” since they “occur in native speech with sufficient consistency to be considered regular features of fluency” (p. 54).

Building on Hieke’s research, Anderson-Hsieh et al. (1994) examined some CSPs, namely linking, flapping, vowel reduction, and deletion, in the English of Japanese ESL learners’ comparing them to NSs of American English. The authors expanded Hieke’s study by examining the production of intermediate-proficiency (IP) and high-proficiency (HP) NNSs, and by exploring the extent to which style shifting affected the CSPs of ESL learners. Results showed that while the HP group approximated the performance of the native speaker group, the IP group

often lagged far behind. An analysis of the reduced forms used revealed that the IP group showed a strong tendency to keep word boundaries intact by inserting a glottal stop before the word-initial vowel in the second word. The HP group showed the same tendency but less frequently. In an effort to examine the minor distinctions in CSPs, Alameen (2007b) replicated Anderson-Hsieh et al.'s (1994) macroanalytical study while focusing on only C-V and V-V linking. Results indicated that beginning-proficiency and intermediate-proficiency participants linked their words significantly less often than NS participants did. However, the linking rates of the two NNS groups were similar despite the difference in proficiency level. While supporting past research findings on linking frequency, results of the study contradicted Anderson-Hsieh et al.'s (1994) results in terms of finding no significant difference between spontaneous and read speech styles. In that study, I was also able to show that native speakers linked more frequently towards function words than to content words.

### **Effectiveness of CSP instruction on L2 production**

Although there have been numerous studies on the effectiveness of teaching CSP on listening perception and comprehension, very little research has been conducted on CSPs production in that regard. This can be largely attributed to the pedagogical priorities of teaching listening to ESL learners since they are more likely to listen than to speak in ESL contexts, and partly to a general belief that CSPs are only a complementary topic in pronunciation teaching and sometimes markers of 'sloppy speech'. As a result, I was able to find only three research studies (Kuo, 2009; Melenca, 2001; Sardegna, 2011) investigating the effectiveness of CSP instruction on L2 learners. Interestingly, all studies were primarily interested in linking, and all were masters or PhD theses. This can probably be accounted for by the facts that (a) linking, especially C-V linking, is the simplest and 'mildest' CSP (Hieke, 1987) since word boundaries are left almost intact, (b) linking as a phenomenon is prevalent in all speech styles, while other CSPs can be found in more informal styles, e.g., palatalization, (c) L2 problems in linking production can render production disconnected and choppy, and hence, difficult for NS to understand (Dauer, 1992) and unlinked speech can sometimes be viewed as aggressive and abrupt (Anderson-Hsieh et al., 1994; Hatch, 1992).

Melenca (2001) explored the influence of explicitly teaching Japanese speakers of English how to connect speech so as to avoid a robotic speech rhythm. A control (N=4) and an experimental group (N=5) were each given three one-hour sessions in English. Their ability to link word pairs was rated using reading aloud and elicited free-speech monologues that were compared to a NS baseline. Descriptive statistics showed that individual performances in pre- and posttest varied considerably. Yet they also demonstrated that the performance of experimental group participants either improved or remained relatively stable in linking ability while the CG performance stayed the same. Noteworthy are the findings that the average percentages of linking while reading a text was at 67% and while speaking freely at 73%. This suggests that linking occurs with approximately equal frequency under both conditions. Melenca, furthermore, recommended that C-V and V-V linking be taught in one type of experiment, while C-C linking should be investigated in a separate study, due to the variety and complexity of C-C linking contexts.

By training EFL elementary school students in Taiwan on features of linking for 14 weeks, Kuo (2009) examined whether such training positively affected students' speech production (see Table 1 for more details on the study). After receiving instruction, the experimental group significantly improved their speech production and developed phonological awareness. Among the taught categories, V-V linking posed more problems for the experimental group due to its high degree of variance.

In spite of the positive influence of training measured immediately after the treatment, effectiveness of the training cannot be fully evaluated without examining the long-term effects of such training. Sardegna (2011) attempted to fill this gap. Using the Covert Rehearsal Model (Dickerson, 1994), she trained 38 international graduate students on how to improve their ability to link sounds within and across words. A read-aloud test was administered and recorded twice during the course, and again five months to two years after the course ended. The results suggested that students maintained a significant improvement over time regardless of their native language, gender, and length of stay in the US prior to instruction. However, other learner characteristics and factors seemed to contribute to greater or lesser improvement over time, namely (a) entering proficiency level with linking, (b) degree of improvement with linking during the course, (c) quantity, quality, and especially frequency of practice with linking in

covert rehearsal, (d) strong intrinsic and extrinsic motivations to improve, and (e) prioritization of linking over other targets for focused practice (p. 116).

More importantly, the previous studies reveal several problem areas on which researchers need to focus in order to optimize time spent in researching linking pronunciation training in particular and CSP training in general:

- (1) A longer period of instruction may facilitate more successful output.
- (2) The reading task approximates the spontaneous speech task in actual linking levels (Alameen, 2007b; Melenca, 2001), in addition to being far easier to measure and control. Therefore, read speech will be utilized for the purposes of this study.
- (3) Practicing all types of linking can be time-consuming and confusing to students (Melenca, 2001). For this reason, only C-V and V-V linking with incidents of /h/-deletion will be addressed in this study.
- (4) There is a need for exploring newer approaches to teaching CSPs that could prove to be beneficial to L2 learners, e.g., the use of electronic visual feedback which will be examined in this study.

### **The Effect of Word Frequency**

The effect of word frequency on language processing has been well-documented (see summary in Ellis, 2002). High frequency words (HF) are words high in number and essential for effective language use, especially in everyday situations. However, the division between HF words and low frequency (LF) words is unclear. In this study, I will use words from the most frequent 1000 word list (K1) for HF words and words from the academic word list for LF words. The academic word list is a corpus-based word list that includes the most frequent words outside the first 2000 most frequently occurring words of English in academic contexts (Coxhead, 2000).

In perception tasks, HF words tend to facilitate word recognition (Bybee, 2001). Kirsner (1994) showed that there are strong effects of word frequency on the speed and accuracy of several word recognition processes, such as speech perception, reading, and object naming; and word production processes, such as speaking, typing, and writing, in both native and non-native speakers. HF words have stronger representations in memory, are more easily accessed and more



likely to undergo reductive sound change. Therefore, HF words may be more easily linked to other words than LF words. The effect of word frequency on NNSs' linking accuracy has not been examined, however. This study investigates whether L2 learners' have more difficulty linking to and from LF words than HF words and whether they improve better in HF contexts after the treatment.

The distribution of HF and LF words is influenced by register. According to Biber (2009), the description of a register has three major components. First, registers are identified for their situational contexts, for instance, whether they are spoken or written and what their communicative purposes are. The second component of register description is concerned with the analysis of the functional associations between linguistic forms and situational/communicative contexts. Third, registers are described for their pervasive linguistic features that include both lexical and grammatical characteristics. Lexical characteristics include both high and low frequency words.

To investigate whether language learners link more to and from HF or LF words, I devised an academic text with a high concentration of LF words, and a short fiction story that includes only K1 HF words. Describing the characteristics of the two registers can help us understand possible variation in learners' ability to link words in them. The primary goal of academic prose is to convey information in certain situations usually connected to academic work, such as university textbooks and academic journals. Therefore, academic prose shows considerable lexical diversity that is reflected in extremely frequent nouns, which are often multisyllabic, low frequency and specialized words (e.g., adjustments, variety). In contrast, verbs are much less common in academic prose than in other registers, although there are specific verb categories that are typical of academic prose (e.g., copula be and passive voice verbs). Although adverbs overall are more common in spoken registers, longer linking adverbs (e.g., however, additionally, etc.) are more common in academic prose.

The purpose of fiction is not to convey information, but rather to tell a story. Linguistically, fiction includes a high frequency of first and third person pronouns, past tense and perfect aspect verbs, adverbials of time, and present participial clauses. Third-person pronouns are used to refer to the participants in the narrative, while past tense and perfect aspects verbs are used to describe the past event that constitute the backbone of the narrative. Present participial

clauses, a type of adverbial subordination, are used to create vivid imagery in the depictive narrative (e.g., John ran out the door, waving cheerfully as he left). Finally, public verbs, i.e., speech act verbs, are used to report the direct and indirect speech acts of the participants (e.g., declare, tell, report, and say).

The distribution of these linguistic features in academic and fiction registers may affect the pronunciation of NSs and NNSs in general and their linking patterns in particular. Examining how learners linked HF and LF words and how their linking improved in each category after training will help us set teaching priorities in the classroom. It will also contribute to our understanding of the relationship between CSP and register variation.

### **Electronic Visual Feedback**

The use of electronic visual feedback (EVF) for pronunciation training goes back to the 1960s with oscilloscopes that displayed a shape corresponding to a signal sustained vowel (James, 1976; Vardanian, 1964). It refers to graphical displays of a speaker's speech signal that are usually represented by spectrograms, waveforms, and pitch tracing, among others. Researchers and teachers have used visual displays to raise learners' awareness and enhance their understanding of a wide variety of segmental and suprasegmental aspects of L2 pronunciation (Anderson-Hsieh, 1992; Chun, 1989; de Bot, 1983; Lane et al., 1988; Molholt, 1988; Pennington & Esling, 1996; Weltens & de Bot, 1984; and more recent work by Coniam, 2002; Hardison, 2004, 2005; Lara, 2009; Levis & Pickering, 2004; Motohashi-Saigo & Hardison, 2009; Ruellot, 2006; Varden, 2006). However, one area rarely explored in EVF literature is the teaching of connected speech processes.

#### **Types of electronic visual feedback**

This section describes three types of EVF: pitch tracings, sound spectrographs, and waveform displays. It discusses the pronunciation problems each one can be used to teach or analyze. Advantages of EVF and limitations on its use are also reported and followed by guidelines for using EVF. The section also reports on studies that have investigated the effectiveness of EVF on L2 pronunciation and surveys students' perspectives on using them for pronunciation training.

### Pitch tracings

Pitch tracings, or visualizations of fundamental frequency (f0) contours, are a technological tool that has appeared ubiquitously in L2 prosody research and pedagogical recommendations. In a pitch contour, pitch is represented on the vertical axis and time on the horizontal axis. Several tools have been employed to display pitch contours of spoken utterances in real time on a computer screen to enable L2 learners to visually compare their speech with a model. Commercially available software such as VisiPitch by Kay-Elementrics (KayPENTAX, 2013) as well as free programs such as PRAAT (Boersma & Weenink, 2013) provide visual representations of pitch contours. Most of the commercial software allow learners to compare their own production to a model utterance. Originally, a dual display on the screen allowed the model utterance to appear on one half of the screen and the students' production attempts on the other half, asking students to compare each attempt with the native speaker model (Anderson-Hsieh, 1992). More recently, the NS's version started to be overlaid on the learner's in a contrasting color to allow for more convenient comparison. In Figure 3 (Hardison, 2004), top screen A is a French native speaker utterance of *Elles adorent la couleur rouge* (They love the color red). The bottom screen B is a learner's production with NS overlay.

For the most part, visual pitch contours have been employed in teaching sentence-level intonation (Hardison, 2004) and discourse-level intonation (Hardison, 2005). They, however, have also been used to teach word-level pitch accents in Japanese (Hirata, 2004), and paratones, that is, initial extra-high pitch levels that mark discourse topic shifts in connected speech (Levis & Pickering, 2004). Several studies have shown that pitch contours can be more effective than auditory-only input for improving learners' intonational patterns in the L2 (e.g., de Bot, 1983; Hardison, 2004; Spaai & Hermes, 1992; Weltens & de Bot, 1984), though no such effects were found in other studies (e.g., Vardanian, 1964).

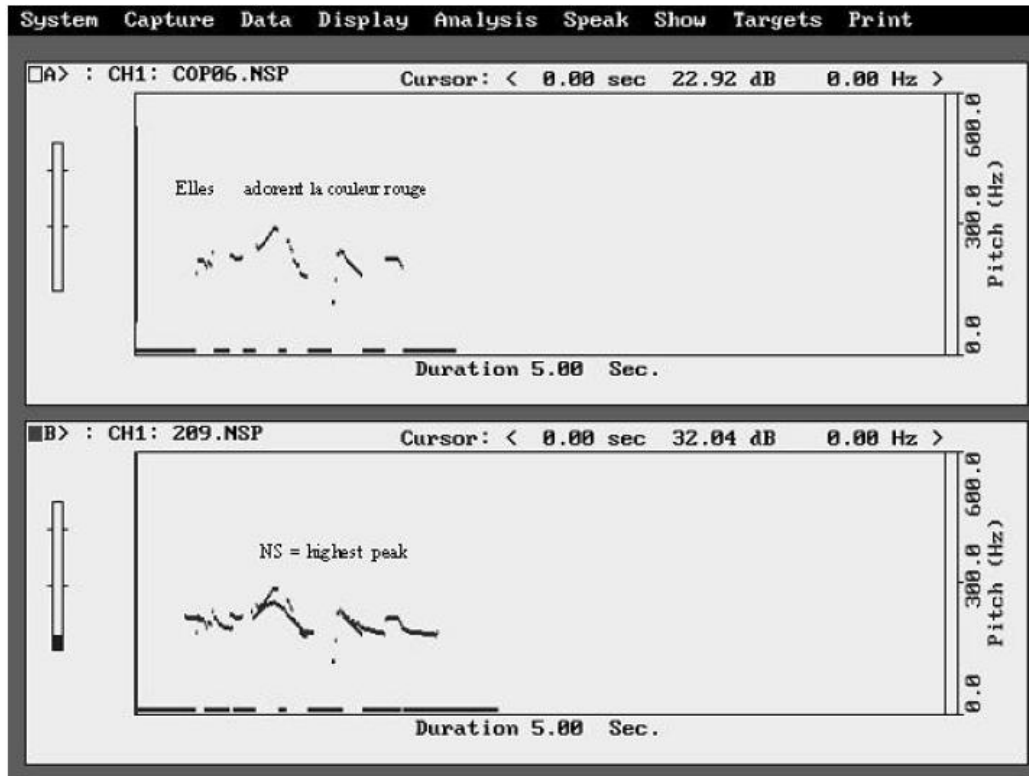


Figure 3. NS pitch contour (top), learner's production (bottom) (Hardison, 2004, p. 43).

### Sound spectrographs

A spectrogram is a “three-dimensional printout of the acoustic properties of the speech signal: time (horizontal), frequency (vertical), and intensity (darkness of the trace)” (Brown, 2013, p. 3). The frequency at the bottom of the spectrogram is usually 0 Hz, and a common value for the frequency at the top is 5000 Hz. Differences in intensity (i.e., amplitude) are shown on a spectrogram by shading. The frequency components with the highest amplitude values are shown in dark black. Components with lower intensity values are displayed in lighter shades of grey to white, with white signifying very low amplitude or silence.

Because of these features, spectrograms have been used in teaching word and sentence stress (Molholt, 1988). Since every phoneme is distinguished by its own unique pattern on the spectrogram, researchers have used spectrograms in teaching certain consonants and vowels. The shape of the vocal tract affects the frequencies at which formants occur. Spectrograms display vowel formants that NNSs produce and compare them to formant measurements of NSs articulation and practice different tongue configurations.

Students are presented with a model spectrogram of a target phoneme or utterance to practice and imitate. However, since spectrograms are representations of raw data, students need training on how to interpret the sound patterns on their screen. Consequently, some researchers were skeptical about their pedagogical value in the L2 classroom (Neri et al., 2002; O'Brien, 2006). The attempts to teach segmentals through speech spectrographic devices have been primarily anecdotal and have concentrated primarily on consonants (aspects such as voicing and aspiration) (Johnson, Dunkel, & Rekart, 1994; Lambacher, 1999; Molholt, 1988) and less often on vowels (Lambacher, 2001; Lara, 2009).

### Waveform displays

The waveform is the most basic display of EVF. It is a graph of sound amplitude (on the vertical axis) against time (on the horizontal axis), “a representation of rapid variations in air pressure caused by actions of a speaker’s vocal organs superimposed on the outgoing airflow” (Motohashi-Saigo & Hardison, 2009, pp. 30–31). The microphone receives sound waves, turning them into an electric signal, which, in turn, is changed into a string of numbers by the computer. The computer software draws those numbers on the screen as strong and weak waves and peaks. The vertical lines in the waveform represent pulses produced by vibrating vocal cords. A waveform of the recording of the utterance *right away* is shown in Figure 1a with the individual words labeled.

Although it is not possible to identify accurately sound boundaries in a waveform, as with a spectrogram, we can identify characteristic features like voicing, occlusions, stop bursts, and fricative hissing. If we analyze the waveform into its frequency components, we obtain a spectrogram which can be deciphered. Figure 4 is an annotated waveform display of *news items*. Notice the voicing at (1) and the silent occlusion of the voiceless stop /t/ at (2) followed by a burst and aspiration at (3).

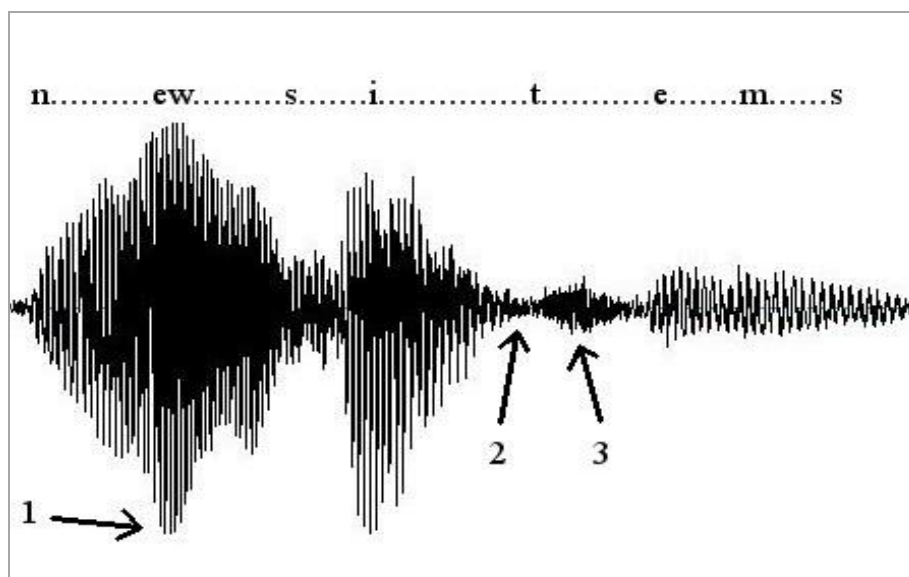


Figure 4. Annotated waveform display of *news items*

### Using waveforms in teaching linking

Waveforms can be useful for practicing English stress, intonation, rhythm and syllables (Anderson-Hsieh, 1992; Varden, 2006). In the current study, waveforms will be used as a tool to raise learners' awareness of linking as a phenomenon in English connected speech since they show the link (or non-link) between words. Waveforms were selected for visual feedback in this study because they (a) easily show the link or no-link between words, (b) are available on many free audio authoring software programs, and (c) are a familiar visual display to most students. In a survey performed during an earlier pilot of this study, 93% of the respondents said they have come across waveforms in music and recording software on their computers and smart phones much more than spectrograms and pitch tracings. Thus, they were more comfortable dealing with a tool they had seen before.

Linking refers to joining the boundary sounds of two words to avoid hiatus, which creates sound continuity distinctive on the waveform. When C-V or V-V linking is appropriately performed, the wave is continuous between the two linked words. However, when the two words are unlinked, the waveform is narrow at the break or sometimes completely interrupted. Figure 5a shows the waveform of two linked words *right away*, while Figure 5b demonstrates the discontinuous link between the two words.

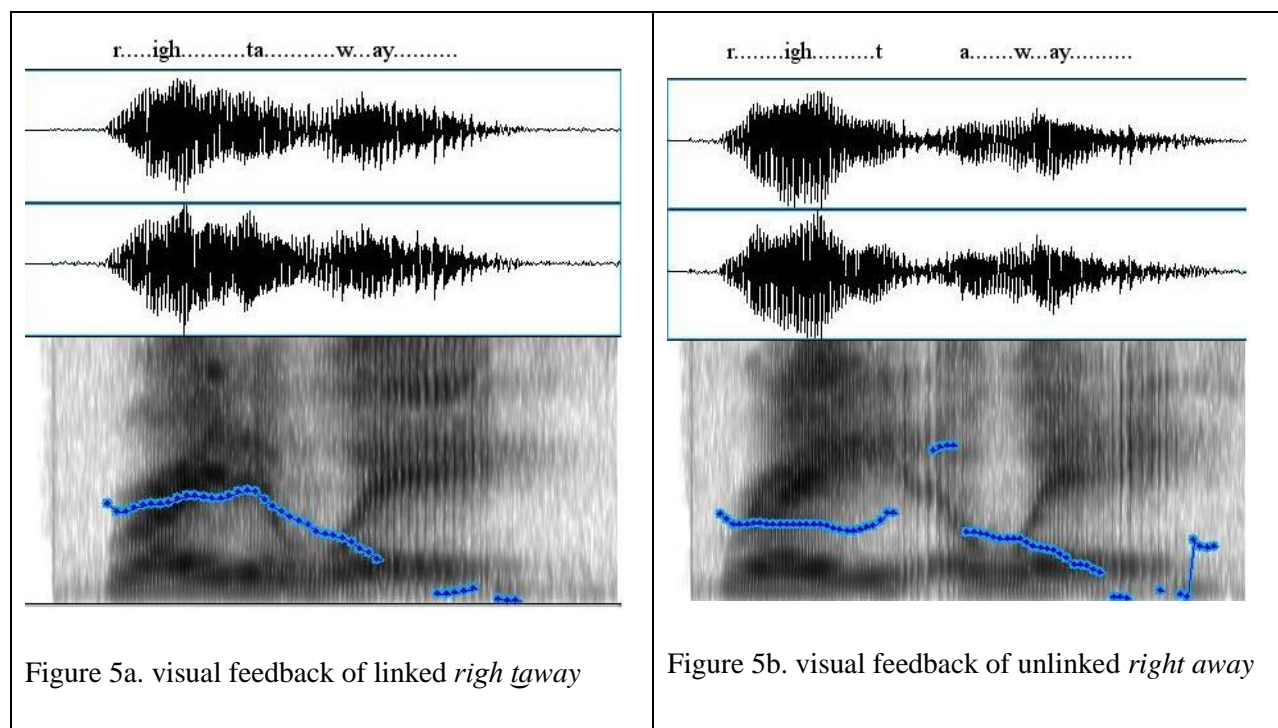


Figure 5. Visual feedback of *right away*

Since spectrograms are used best for segmental training and since students need significant amount of training before they are able to interpret them, I did not include them as visual feedback to practice linking. Pitch contours, on the other hand can sometimes display lines of continuity and interruption, indicating linking; however, they also contain other irrelevant microintonation that could be confusing to learners. All types of EVF equally cannot display voiceless consonants, so students and teachers should be made aware of this.

Figure 6 displays another example of the difference between linked and unlinked utterances with V-V linking. While the waveform of unlinked 'say it' has two distinctive waves for each word with a short stop in between them, the waveform of the linked words is one longer wave with no pause between the words.

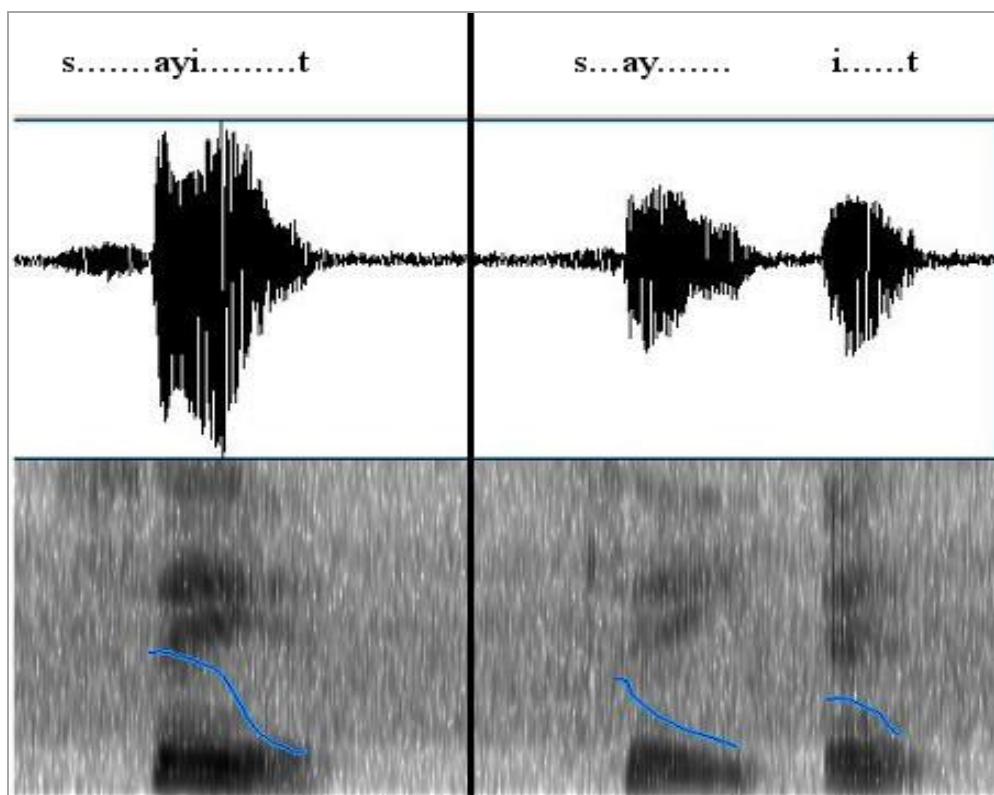


Figure 6. Visual feedback of *say it*

### **EVF software**

Software with EVF tools has become increasingly accessible and affordable, enabling systematic research in the teaching of prosody using these tools and classroom application. Some systems provide instantaneous EVF that makes possible a direct comparison of learners' utterances with stored utterances or displays recorded by the teacher or a native speaker. Most of these systems are generally authoring tools made primarily to support phoneticians and speech scientists in specialized scientific research. However, many have been adapted to be used by applied linguists, language teachers and learners. Notable systems are VisiPitch by Kay-Elmetrics (KayPENTAX, 2013; Molholt, 1988) and WinPitchLTL by Pitch Instruments Inc. (2013). Due to technological requirements and/or program costs, few EVF systems have a real chance to being adopted by L2 programs. Nevertheless, for those teachers and students interested in using EVF but lack appropriate technological expertise and/or money resources, Praat (Boersma & Weenink, 2013) and Audacity (Audacity, 2013) are good alternatives. They are free, open-source audio-authoring programs compatible with several operating systems.



Audacity, the software used in the training of the present study, is an easy-to-use audio editor and recorder developed by a group of volunteers and distributed under the GNU General Public License (GPL). It has a friendly interface, accessible user manuals and support groups which facilitate its incorporation into teacher training programs and course materials. Teachers and students can use Audacity for a range of language learning and teaching tasks, such as listening, speaking, and pronunciation practice (Alameen, 2007a). It allows users to view, edit, and manipulate any audio track as a spectrogram or a waveform. Users can, furthermore, open several Audacity windows to compare the visual displays of multiple utterances. Audacity was the software of choice for the current study because of its ease-of-use, waveform features, stability, and availability in all language labs.

### **Advantages and limitations of EVF**

The primary purpose of implementing visual displays in listening and pronunciation instruction is to provide learners with an additional source of feedback that might help them “see” what they have difficulty perceiving when they receive audio feedback only. Learners develop specialized abilities to perceive their native language, but they still need to develop sensitivity to L2 speech. This is what EVF can help with (de Bot, 1983; Varden, 2006). In addition, research in computer-assisted pronunciation learning has demonstrated that learners who receive both audio and visual feedback perform significantly better than learners who receive audio feedback only (de Bot, 1983; Hardison, 2004; Motohashi-Saigo & Hardison, 2009). Information processed through more than one cognitive channel is likely to be retrieved more successfully than when it is processed through only one channel (Paivio, 1971, 1991). This means that learners in the AV group may outperform those in the AO group since they benefit from feedback that was presented auditorily, but also visually and verbally.

Visual displays can make suprasegmentals tangible to students and provide a convenient means for analyzing and discussing the students’ problems (Anderson-Hsieh, 1992). Learners can, furthermore, compare their own pronunciation with that of a NS model. This helps them recognize the type and the significance of the error being produced, provided they receive appropriate training on how to do it. The automatic and objective feedback of visual displays allows them to self-monitor and self-evaluate their production. This individualization of input

and feedback enables learners to have greater control over their learning process, which is something difficult to accomplish through traditional pronunciation approaches (Lara, 2009). Finally, EVF is a quick, repeatable, reliable, precise, salient, and objective method of presenting authentic and nonauthentic speech to L2 learners (Pennington, 1999).

Despite the many advantages they potentially offer, EVF systems present limitations that need to be addressed if learners are to fully benefit from this technology. First, the richness of information presented in EVF can be overwhelming to users who may not have the knowledge to interpret such representation of raw data (Neri et al., 2002). If learners cannot decipher these displays and extract meaningful feedback, they are likely to make random attempts at correcting the presumed errors which, instead of improving pronunciation, may have the effect of reinforcing poor pronunciation (Eskenazi, 1999). Second, the learners might develop inaccurate assumptions about replicating visual displays (Neri et al., 2002). The fact that the system shows two comparable displays of the NNS learner and the NS model wrongly suggests that the ultimate aim of training is to produce an utterance whose visual displays closely corresponds to that of the model utterance. As a matter of fact, two well-pronounced utterances with the same content may have waveforms or spectrograms that are different from each other. Third, EVF systems lack baseline differentiating “variation from (true) error” (Pennington, 1999, p. 431). The lack of baseline entails the risk for the learner of receiving “false negative” feedback indicating that the learner’s production was not correct when in fact it is, and “false positive” feedback indicating that the production is correct when it is not within the range of acceptable performance. Finally, certain aspects of pronunciation do not show up well in the visual representations of the speech analysis such as waveforms and pitch contours. Because only voiced sounds are represented in the pitch displays, their quality is dependent on the proportion of voiced sounds that appear in the utterance being displayed (Anderson-Hsieh, 1994; Pennington, 1999).

### **Guidelines for using EVF**

The visual displays that learners see on the screen serve as templates that will aid them in learning to recognize and produce patterns of connected speech. They should, therefore, be as clear and easily interpretable as possible. There are a number of guidelines that researchers,

developers, and teachers can follow to control the effects of the aforementioned EFV limitations. Realizing the limitations of the medium, the present study will attempt to implement the following guidelines.

In order to help L2 learners interpret the rich information presented in visual displays, waveforms and spectrograms will be annotated in a way that highlights the relationship between the auditory signal and its visual representation. For example, a print screen of a waveform can be modified by adding arrows, lines, colors, and text demonstrating where each sound/word begins and ends. Such modification will make salient target features of pronunciation. Figure 7 shows the waveform of ‘he’s a great cook’ where C-V linking between ‘he’s’ and ‘a’ is highlighted by annotation.

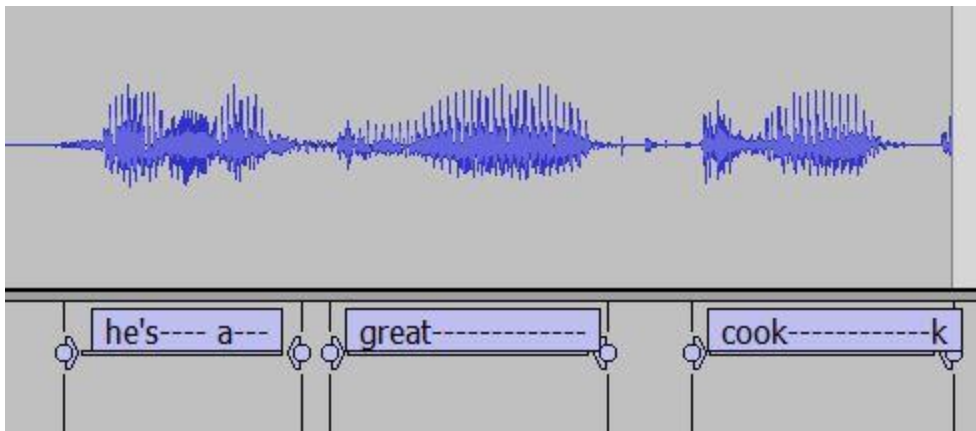


Figure 7. Waveform of *he's a great cook* showing linking between *he's* and *a*

Moreover, learners should be made aware of the fact that visual displays of two similar utterances can differ and that EVF should be primarily used as a tool to raise learner's awareness of what happens at word boundaries (Anderson-Hsieh, 1992). For this purpose, a training plan should be implemented at the beginning of instruction to guide the learner in the acquisition of skills enabling the exploitation of the information provided by the EVF system. After reviewing 17 programs for improving pronunciation, Verhofstadt (2002) concluded that speech visualization systems show pedagogical promise “provided that the phonetic material is carefully selected so that it best serves the didactic goal, and that there is enough explicit phonological advice and theoretical guidance about how to interpret the display” (p. 182). Indeed, the effectiveness of EVF depends a great deal on the teacher's efforts in planning student activities

as well as his or her ability to both understand the results of phonetic analyses performed and to impart this understanding to the learners in the class. In addition, the teacher should be willing to devote time to create appropriate training and exercises for students with EVF which, in many cases, is not practical (O'Brien, 2006).

Capitalizing on these guidelines, the training procedure of the present investigation includes basic training in waveform reading. In this training, L2 learners in the AV group were taught (a) what a waveform is, (b) how to record audio using Audacity and generate a waveform, (c) how to identify target characteristics of a waveform, (d) how to identify visual differences between the linked and unlinked forms, and (e) how voiceless sounds are represented in waveforms. Such training not only helps learners decipher waveforms correctly, but also enables them to monitor themselves for correct use of linking.

### **Effectiveness of EVF on L2 pronunciation**

A number of studies have focused on the use of EVF for the teaching of segmental aspects in pronunciation such as high front vowels by pre-service NN English teachers (Lara, 2009), L2 Japanese geminates by beginning-level L1 English learners (Motohashi-Saigo & Hardison, 2009); and French segmental pronunciation at the beginner level (Ruellot, 2006). Nonetheless, the bulk of EVF research has concentrated on the instruction of L2 suprasegmental features, primarily teaching intonation (Anderson-Hsieh, 1996; Chun, 1989; de Bot, 1983; Hardison, 2004; Levis & Pickering, 2004; Molholt, 1988; Spaai & Hermes, 1992; Vardanian, 1964; Weltens & de Bot, 1984). These studies have shown that the use of pitch contours contributes to students' success in achieving improved intonational patterns in the L2. For example, Hardison (2004) showed that after three weeks of prosodic training using pitch contours, learners of French improved significantly in prosody with generalization to segmental production and novel sentences.

Additionally, other studies described, in more detail, the implementation of EVF systems in L2 classrooms to teach aspects such as stress, rhythm, and connected speech (Anderson-Hsieh, 1992, 1994; Coniam, 2002; Varden, 2006). It is to be noted that Varden's study is the only one in the literature that explores the use of EVF in pronunciation training with a focus on CSPs, namely schwa reduction, flapping, coalescent assimilation, and elision. The author provides

extensive explanations for the use of WASP graphs (a free phonetic software package, <http://www.phon.ucl.ac.uk/resource/sfs/wasp.htm>) of reductions and warnings of the difficulties that can be encountered in using such software for pronunciation training. In terms of visual display types used, while most segmental studies have explored the use of spectrograms, researchers investigating suprasegmental aspects have mainly used pitch contours. Only two studies explored the use of waveforms in teaching stress timing and rhythm (Coniam, 2002), and Japanese geminates (Motohashi-Saigo & Hardison, 2009).

Coniam (2002) used waveforms to raise language teachers' awareness of the stress-timed nature of American English in relation to the English variety spoken in Hong Kong where the experiment took place. He captured and annotated audio waveforms extracted from a popular TV show in Hong Kong whose main characters spoke both Hong Kong and American English. The teachers viewed the one-minute clip and discussed 'wooden' speech patterns corresponding to Hong Kong English. They received basic training on how to interpret waveforms in terms of how high they appear to 'peak', and the relative length of certain syllables compared to others. Coniam drew arrows to relate the major stressed syllables in the speakers' utterances to the energy 'peaks' in the waveforms. It was possible to visually differentiate between waveforms from the two varieties since the NNSs had double the number of energy peaks than the NSs in the clips. This approach was successful in helping teachers understand this difference in the two varieties mainly because Coniam understood the limitations of waveforms and capitalized on its strengths by "focusing only on the relatively iconic visual displays of syllable energy while also providing a significant amount of scaffolding that identified the location of the words associated with each peak" (Levis, 2007, p. 191).

Participants in Motohashi-Saigo and Hardison's (2009) study were asked to complete 10 self-paced training sessions outside class. Speech waveform displays were chosen to emphasize segmental duration of Japanese geminates as they are perceived by American English learners. The training included a forced-choice identification task that asked learners to listen to a stimulus (audio only group) or listen and watch the waveform of the stimulus (audio visual group) then choose one of three minimal triplets: singleton (e.g., *sasu*), geminate (e.g., *sassu*), and long vowel (e.g., *saasu*). Results revealed that mean scores improved the most for the AV group although

there was no statistically significant effect of condition. Test of generalization revealed a 5% decline in accuracy for AV and 14% for A-only.

### **Student perspectives on using EVF**

Researchers also administered qualitative methods (such as questionnaires, interviews, and class observation) to participants in the experimental group(s) to gather their impressions regarding the use of EVF in teaching the target pronunciation aspects, and to gain insight into whether EVF can be successfully implemented in the L2 classroom. Most participants thought that EVF was an effective approach to improving perceptual awareness and pronunciation, and had the potential to be successfully implemented in the second language classroom (Anderson-Hsieh, 1992; Coniam, 2002; Hardison, 2004; Lara, 2009; Motohashi-Saigo & Hardison, 2009; Ruellot, 2006). However, participants in Ruellot (2006) found audio feedback to be more helpful than both the visual feedback and the verbal comments accompanying it.

The positive participant feedback highlighted several issues. Incorporating EVF into L2 instruction brought about greater awareness of segmental and suprasegmental aspects of speech. Visual displays showed learners where to focus their attention and gave them a perspective that they had not been able to appreciate before (Coniam, 2002; Motohashi-Saigo & Hardison, 2009). They could see how their and other NNS utterances sounded different from the NS ones in terms of intonation, stress, and syllable timing. Moreover, it was believed that EVF increased learner's confidence in producing L2 utterances after prosodic training using pitch contours (Hardison, 2004).

On the other hand, some learners and nonnative L2 teachers believed that the effectiveness and implementation of EVF instruction could be greatly improved if special attention was given to planning, learner and teacher training, and accessibility. In Ruellot's (2006) dissertation, French learners suggested adding more elaborate explanations on how to interpret the line patterns on spectrograms, and how to approximate the NS model's pattern and pronunciation. Due to the online autonomous nature of the activity, the participants saw the need for the presence of a human expert who could initially train them in interpreting EVF and help answer any questions. Benefits of visual feedback could also be maximized by allocating more time for the learners to process the visual feedback (Lara, 2009; Ruellot, 2006). It was also felt

that applying these suggestions in conjunction with traditional pronunciation instructional procedures could enhance student's learning experience.

From the studies reviewed, there is a need to conduct more empirical research on the effectiveness of EVF in the acquisition of connected speech processes. As previously mentioned, the benefits of using visual displays to teach suprasegmental aspects of pronunciation are worth the effort provided that EVF limitations are appropriately addressed. Therefore, the current study tests the effectiveness of instruction using audio-only or audio-visual feedback (i. e., waveforms) on the acquisition of the English C-C and V-V linking by ESL students in the US. It also strives to determine how effective learners perceived these types of training to be, and how to improve this instructional method for future implementation.

## **Chapter Summary**

This chapter reviewed the literature that provides a background for this dissertation work. It provided an overview of connected speech processes and discussed their classification, features, and the linguistics and social factors that affect them in speech. Then, linking was defined and its three types introduced. The next section examined strategies NSs and NNSs use to segment connected speech, highlighted the effect of CSPs on L2 listening and reviewed the literature on the effectiveness of CSPs perceptual training on listening perception and comprehension. This was followed by a parallel overview of literature on the production of CSPs and the effectiveness of instruction on pronouncing them. Examination of previous research generally indicated that CSP instruction facilitated learners' perception and production of connected speech. However, most studies failed to address the long-term effects of such training on learners' perceptual accuracy. Moreover, no study has investigated generalization and transfer of improvement to novel contexts which indicates that improved abilities could extend beyond the training to natural language usage.

The chapter also included a discussion of the effect of word frequency on linking and detailed differences between registers that can lead to variation in linking patterns. The final part of the chapter focused on EVF and its use in teaching and researching pronunciation. The information introduced suggests that waveform displays are an appropriate form of visual feedback to use in teaching linking when certain guidelines are followed.

## **CHAPTER III: METHODS**

This chapter describes the methodology undertaken in this study, which followed a quasi-experimental design to evaluate the effectiveness of two training approaches, audio-only and audio-visual feedback, on learners' linking perception and production. Quantitative and qualitative data were collected from 44 ESL learners to answer the research questions. After describing the research design, the characteristics of the participants are enlisted together with details about the setting. The data collection procedure and study timeline are provided next. Following that, a detailed account of the training materials is provided. It includes information about the procedure of creating the online activities, stimuli development guidelines, and examples of AO and AV training materials. The quantitative and qualitative data collection tools and analysis procedure as well as results and discussion are detailed in the following three chapters.

### **Research Design**

The study used a quasi-experimental design with pretest, posttest, delayed posttest, and a control group. Quasi-experimental design is the same as the standard controlled experimental design except that participants are not randomly assigned to the experimental and control groups (Mackey & Gass, 2005). Since the study took place in intact ESL classes, random assignment of students by the researcher was not possible; therefore, the quasi-experimental design was implemented. Because of such practical constraints, working with 'non-equivalent groups' has become an accepted research methodology in applied linguistics studies where randomization is impossible or impractical (Dörnyei, 2007). In order to improve the design of the study, students were not allowed to self-select to be in a treatment group or a control group (Heinsman & Shadish, 1996). Furthermore, Ortega and Iberri-Shea (2005) argued that the quality of quasi-experimental design has shown considerable improvement in recent years, featuring longer treatment periods and immediate as well as delayed posttests. For this reason, in addition to the posttest administered immediately after treatment, this study also included a delayed posttest taken three weeks after treatment.



In addition, the current study examined the efficacy of AO and AV training in linking to novel contexts. Generalization and transfer of improvement has emerged over the past 20 years as an indicator of successful perception training. Such transfer entails that there is a potential for improved perception and production abilities to extend beyond the training to the natural language usage, which is the ultimate learning objective (Hardison, 2003). A number of studies have demonstrated successful L2 perception and/or production training for certain segmental and pitch features and have been characterized by findings that support the ability of learners to generalize improved perceptual capabilities to novel stimuli (e.g., Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Hardison, 2003; Hirata, 2004; Lively, Logan & Pisoni, 1993; Motohashi-Saigo & Hardison, 2009; Ruellot, 2006). Consequently, learners' performance in linking perception and production in this study were assessed in novel unpracticed contexts and compared to old practiced ones.

The quantitative phase of the study was followed by a qualitative component to add depth to the quantitative results. This mixed-methods approach provides a broader view of the topic and sheds light on findings and the issues investigated from a variety of perspectives (Dörnyei, 2007). Survey questions elicited responses from participants regarding their experiences using the linking training website and the two types of feedback.

The research design was developed in a way that helps to answer the following research questions:

1. How are AO and AV training effective in improving linking perception and production for non-native learners of English?
2. How is the improvement caused by AO and AV training sustained over time?
3. How do AO and AV training generalize to novel perception and production contexts?
4. What is the impact of word frequency on linking production? Are lower frequency words more difficult to link to and from than higher frequency words? Do learners improve better after treatment in HF contexts than in LF contexts?
5. What are learners' perceptions of the use of audio-only and electronic visual feedback in teaching English linking?

## Participants and Setting Description

A questionnaire was distributed before the training to collect background data from participants about their former and current EFL/ESL experiences, especially those related to pronunciation. Nicknames replaced participants' real names to protect their identities (see Appendix I for Pre-Training Questionnaire).

Participants of the research study were recruited from Engl 99L classes: Strategies for Listening at Iowa State University. This required course allows ESL students the opportunity to develop their academic listening skills in preparation for their study in the US. Most students take these courses during their first year. The study took place in three intact classes of Engl 99L where every section was randomly assigned a type of treatment: the control group (CG) received no treatment, the AO group received training with audio-only feedback, and the audio-visual (AV) group received training with electronic audio-visual feedback.

The majority of enrolled students were undergraduates with a small percentage of graduate students. In addition, the participants included a large number of Chinese native speaking students due to the high proportion of Chinese students in the international student population of Iowa State University (54.6% of the international student population) (Iowa State University Office of Institutional Research, 2013). Participants with prior training in connected speech perception and/or production were excluded from the study. In addition, participants who missed more than one training session, the pretest, or the posttest were not included in the study. Furthermore, participants' proficiency level in English listening was evaluated at the beginning of the semester as a part of a diagnostic test. The test evaluated students' overall listening comprehension and their perception of connected speech. Table 3 presents the mean and standard deviation (SD) for the three groups in the listening proficiency test scores as well as the mean and SD for the connected speech section scores of that test.

A total of 45 NNSs students participated in this study with 15 students in every one of the three groups: CG, AO, and AV. The control group participated in the pre- and posttests but did not receive any training. For the production part of the study, three CG participants were excluded from the study because the audio files of their posttest were corrupted. Therefore, the production part had 42 participants. The students were of varied educational backgrounds and

academic majors. Most participants were undergraduate students and six were graduate students. They were from different academic fields such as accounting, computer science, civil engineering, biology, and the like. Their ages ranged from 18-27 with a median of 20. They were 26 males and 18 females. Most of them spoke Chinese as the home language (32), and the others spoke Korean (5), Malay (2), Spanish (1), Hindi (1), Portuguese (1), Cantonese (1) and Arabic (1). More than half of the students in this study were taking the listening class during the first year of their undergraduate study (N= 27). The remaining had been in the US between one and two years. Only twelve students said they had received pronunciation instruction before they came to the US, and nine students said they had received pronunciation instruction since they arrived in the US. However, none of the students had training in CSPs (two students were originally excluded from the study because of they received training on CSP production in the US). An overview of these students can be seen in Table 2.

Table 2. Participant Overview

<b>Variable Categories</b>	<b>Participant Variable</b>	<b>CG</b>	<b>AO</b>	<b>AV</b>
Gender	Male	9	10	7
	Female	6	5	8
Native language	Chinese	11	11	10
	Korean	2	3	1
	Other	2	1	4
Time in the US <sup>a</sup>	Mean	9.93	12.40	4.47
	SD	9.64	11.99	4.20
Listening test score <sup>b</sup>	Mean	40.71	49.47	47.20
	SD	12.09	13.42	14.47
Listening test (CS section)	Mean	39.64	45	41.02
	SD	13.79	14.26	12.89

Note. <sup>a</sup>= time is measured in months; <sup>b</sup>= Maximum score for listening test is 100

Engl 99L sections are usually taught by different instructors who are teaching assistants and PhD students in the Applied Linguistics program. Their course textbook and curricula focus mainly on teaching top-down listening strategies that help learners comprehend text- and sentence-level language; with very little reference to training on bottom-up listening strategies

that help learners focus on the word level of language (perception of linking is a bottom-up listening strategy). To control for instructor teaching style and experience effect, I taught the experimental groups during training. The instructors had the choice of observing or leaving the classroom and both chose to leave. The training was incorporated in the syllabus by the instructors of the experimental groups from the beginning of the semester. All students were asked to participate in the training as a class activity, but only the results of those signed the consent forms were retained.

## **Procedures**

After I announced the study and explained it to the potential participants in Engl 99L classes, I answered their questions. Then, interested students signed the consent form provided. The consent form explained the purpose of the study and training and outlined the tasks, the risks and the benefits (see Appendix II. Consent Form). The training started one week afterward and extended over a period of two weeks. Students received four sessions of training of 50 minutes each. One group had their training sessions on Monday and Wednesday and the other group on Tuesday and Thursday. All sessions took place in a computer lab where students could use headphones with attached microphones. Students signed in to the training Moodle for access to the materials and perception tests. One week after introducing the study, they started their first training session in class by completing the background questionnaire. After that, they took the perception pretest in the lab and the production pretest in a different room. Figure 8 outlines the study's procedure.

Every student met separately with a trained assistant in an isolated room to record the production pretest. After having the procedure explained and their questions answered, students read the production stimuli twice. Only the second reading was considered for analysis purposes to make certain that any pauses or hesitation made while reading were not due to novelty of the material. The participants were asked to read the text aloud in their normal reading style and pause at the red slashes. Speech was recorded using digital audio recorders at a 44,100 Hz sampling rate.

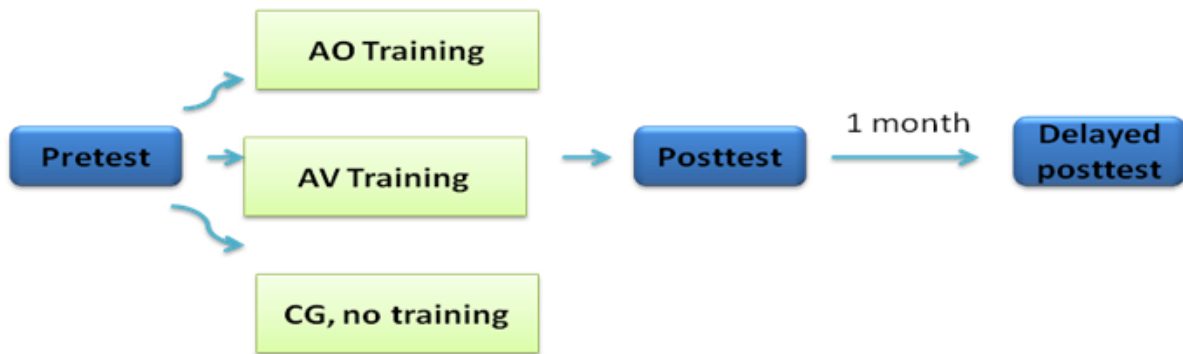


Figure 8. Outline of the study procedure

After the pretest, AO and AV participants worked on Day 1 training materials. CG participants worked on a listening comprehension activity not related to the training. They were offered similar training at the end of the study. Since none of the Engl 99L instructors was experienced in teaching linking and using EVF, I conducted all class activities. This included giving presentations, supervising tests and training, answering students' questions, directing them on how to perform certain tasks, and dealing with technical difficulties.

For both AO and AV groups, day 2 started with a class presentation and practice about linking followed by online training. Days 3 and 4 included shorter presentations and practice and additional online training. Students were not allowed to work on the training materials outside class to control for time of training. On Day 5, students worked on a 20-minute training session, and then took the posttest and post-training questionnaire. The posttest and the questionnaire followed the same procedure as the pretest and pre-training questionnaire. The delayed posttest took place one month after the posttest and followed the same procedure of the other tests.

### Training materials

Over a period of two weeks, the teaching unit activities covered the whole class time for the AO and AV groups. Participants from the two experimental groups worked on in-class and online materials embedded in Moodle. Moodle is a course management system that can integrate audio and video into a variety of activities. The students also used voice authoring tools (Audacity and QuickTime) to record their speech production and practice linking.

The design of the training units is guided by Chapelle's (1998) criteria for developing multimedia CALL. Linking pairs were made salient on the screen either using Moodle question features or waveform annotations. The input mode was changed from text to audio (AO group) or from text to audio and waveforms (AV group) to provide students with a modified reception method. Such materials "hold the potential to provide learners with comprehensible input rather than just input" (Chapelle, 1998, p. 27). Furthermore, the online units provided opportunities for learners to notice their errors in several ways. The majority of perception activities provided immediate feedback upon completion. However, production feedback was mainly learner-centered where the students self-monitored their own utterances and compared them to NS models with or without waveforms. AO activities provided only audio feedback to students. Students listened and compared their pronunciation to that of a model native speaker. In contrast, AV activities presented audio and waveform feedback to students. This, in turn, provided opportunities for learners to correct their linguistic output. In addition, students had the freedom to redo any activity since scores were not going to be tracked during training. Students could work at their own pace; those who needed more time and practice could retake an activity, while those who progressed faster could work on other supplemental activities created for this purpose.

For the AV group, EVF was used for both production and perception activities to raise learners' awareness of linking between words. In addition, participants were able to compare the EVF of their production and that of a native speaker model. It is to be noted, though, that students were not encouraged to produce an identical waveform to that of the model. Only short utterances (2-6 words) were supported by waveforms. To help students understand waveform structure, I manually annotated them in Audacity by adding text that synched with the waveform audio to show where every word begins and ends (see Figure 10 below). I also explained the waveforms in terms of how they are connected, which essentially indicates linking or unlinking, while avoiding technical terms.

Stimuli used in the activities were either recorded by NS speakers, taken from pronunciation textbook listening materials or from online resources. The following measures were taken into account when selecting and creating the training stimuli:

- The stimuli were appropriate for the level of intermediate learners of English in terms of vocabulary choice and syntactic complexity (Chapelle, 2001). Utterance length increased

gradually as students progressed in the training session. As noted by Skehan (1998), material which is both linguistically and conceptually challenging can provide too great a cognitive load for effective processing.

- Sentences were checked for linking frequency. In addition, links with voiceless consonants were avoided as much as possible, although the participants were made aware of the issue of voiceless consonants in EVF during training.
- Speakers in the audio recordings were from a variety of ethnic and cultural backgrounds, especially in the authentic texts.
- Materials were recycled in different exercises. Gapped dictation of paragraphs, for example, were followed by more bottom-up listening and speaking activities, such as 'Read and Compare'. Such activities help students practice listening to and producing shorter segments taken from the previous paragraph.
- Materials included a range of nonauthentic and authentic utterances and offered a gradual transition from the former to the latter. Nonauthentic materials were taken from ESL textbooks. They were initially created for teaching certain pronunciation and listening aspects (not necessarily connected speech). Authentic materials were taken from popular American movies and are divided into shorter passages which, in turn, are divided into shorter utterances. The use of video and film, radio broadcasts, and television programs involves students in activities that present real life listening contexts (Herron & Seay, 1991). According to Omaggio Hadley (1993) and Rogers and Medley (1988), the use of authentic materials in the classroom helps students develop a functional proficiency in the language and use the language communicatively in the real world. Omaggio Hadley (1993) advocates the need for a blend of authentic materials and materials created for instructional purposes.

The study adopted an explicit approach to teaching since this was found to be an effective approach in improving NNSs' perception and production of connected speech (e.g., Carreira, 2008; Crawford, 2006; Kuo, 2009; Matsuzawa, 2006; Melenca, 2001; Wang, 2005). Students learned about linking from a longer presentation on the first day of training and shorter ones at the beginning of the following two days. They had the opportunity to work collaboratively on sample activities during these times, before the individual training online. Additionally,

participants were constantly engaged in a dialog to report their observations about linking. They followed a guided path of exercises that aimed to improve their perception and/or production of linking. A complete list and description of the training unit activities can be found in Table 3.

Table 3. Training unit activities

Activity	Description	Objective	Examples
<b>Counting</b>	Students listened to short sentences and decided how many words there were in every one of them. They were reminded that “small words like 'a, an, of' are counted as words”.	Perception practice: Students were introduced to the concept of linking and develop their ability to recognize linked words.	Should I call him up?
<b>Feedback</b>	Students answered short questions twice in every training session. On the first day, they were asked about what they had noticed while working on the activities regarding connecting words in English. Later on, they were asked to provide short feedback on the part of the training sessions they had finished and report their observations.	These were meant to provide constant source of feedback as well as instantly capture caveats in material design or learner difficulties with linking or the training before the next training session.	- What difficulties have you have while working on the previous questions?  - What helped you answer the questions?
<b>Linked or Not</b>	AO students listened to every utterance (not necessarily a complete sentence) and decided whether the words are linked together or unlinked. On the other hand, AV students listened to every utterance and chose the matching waveform. They were reminded that “when the words are connected, their waveform is also more connected.”	Perception practice: students developed their ability to recognize linked words. Furthermore, AV students practiced reading waveforms and using them as indicators of linked and unlinked utterances.	She asked for it.



Table 3 (Continued)

Activity	Description	Objective	Examples
<b>Mark the Link and Read</b>	<p>AO students listened to every utterance and decided what words are linked together. Then, using QuickTime, audio recording software that does not have EVF features, they practiced these utterances by recording and comparing them to the NS's recording (Figure 9).</p> <p>In contrast, AV students listened to every utterance while following a screenshot of their annotated waveforms. They decided what words are linked. Using Audacity, audio recording software that has EVF features, they practiced by recording these utterances and comparing their audio and waveform to the NSs'. They were reminded that their "waveform should be connected in the same way" (Figure 10).</p>	<p>Perception and production practice: Students developed their ability to recognize linked words. In addition, they practiced linking production by attempting to approximate NSs' utterances and waveforms.</p>	<p>- Is (*) it black (*) or white?</p> <p>- I think you've (*) improved (*) a great deal.</p> <p>(The asterisks indicate the links students are asked to examine.)</p>
<b>Gapped dictation</b>	<p>Students listened to longer audio segments and filled in every blank with two missing words. They were reminded that the two missing words are linked, so they needed to focus on the beginning and the end of the missing utterance.</p>	<p>Perception practice: Students developed their ability to recognize linked words in longer segments of speech.</p>	<p>Short to medium-length paragraphs (Figure 11)</p>
<b>Read and Compare</b>	<p>AO students read the utterances aloud several times paying attention to the links between words. Then, they recorded them using QuickTime and compared their recording to the audio on the website. Students were encouraged to keep practicing until they were able to connect their words like the speaker in the audio. They continued utterance when they were satisfied with their pronunciation.</p>	<p>Production practice: Students practiced linking by attempting to approximate NSs' utterances and waveforms. All utterances in this activity were taken from the previous one 'gapped dictation'.</p>	<p>- They'd checked on him at eight o'clock.</p> <p>- So I studied a lot of languages.</p>

Table 3 (Continued)

	AV students, on the other hand, were asked to record themselves using Audacity and to compare both their audio and their waveform to the one on the website. They were encouraged to keep practicing until they were able to connect their words like the speaker in the audio and until their waveform was connected like the one on the website.	This is meant to provide further focused practice on the previously-practiced longer segment and not to overwhelm students with long text.	
--	--	--	--

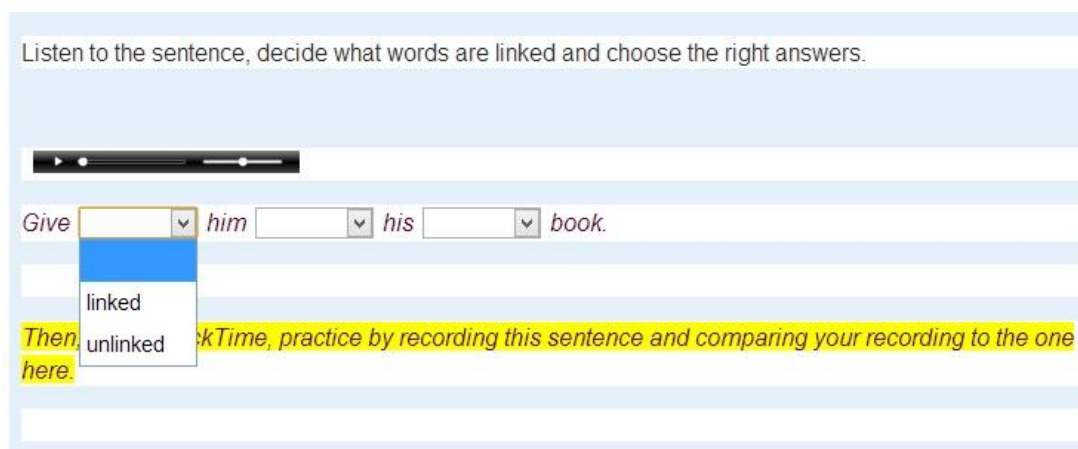

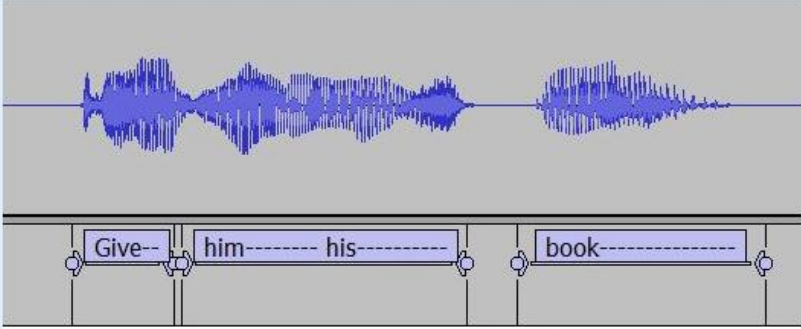


Figure 9. Mark the link activity – AO group

Listen to the sentence and look at its waveform, then decide what words are linked and choose the right answers.



Give   him  his  book.



*Then, using Audacity, practice by recording this sentence and comparing your waveform with the one here. Your waveform should be connected in the same way.*

Figure 10. Mark the link activity – AV group

### In-class presentations

The purpose of the in-class presentations was to provide the participants with background about linking and help to navigate the training unit. The presentations were identical for the AO and the AV group, except for the section on EVF. The AV group presentation had extra slides that introduced waveforms and a different set of activities that incorporated waveforms. Their presentations were consequently longer than the AO's group. Every presentation started with discussing specific aspects of linking with examples. This was followed by exercises identical to the ones used in the training unit on which students worked independently then compared their answers with a classmate. Practicing with sample questions was intended to help familiar students with linking and questions types before students start working on the online training unit autonomously.

Listen to the audio and fill in the blanks with the **TWO missing words**. Remember that the two missing words are linked, so focus at the beginning and the end of the missing utterance.



Zack works downtown, [ ] lives in the suburbs and commutes. He can get to the [ ] lots of ways. Sometime he drives, [ ]'s not worth the trouble. There's too much [ ] rush hour, and it's too expensive to [ ] car. So he catches the bus at the corner and [ ] downtown. When he travels for business [ ] Monday, he [ ] on Sunday because he can get to the airport in forty minutes. The same trip takes [ ] hour on Monday.



Debbie: "I can't help you."

Larry: "Uh, Debbie? [ ] call you Debbie? 'Cause, uh, I [ ] connection [ ] entered this office. And I don't know, I feel like you did too."

Debbie: "I didn't [ ] connection."

(From the movie *Night at the Museum*)

Figure 11. Gapped dictation activity (same for both AO and AV groups)

Day 2 presentation lasted between 20-30 minutes and focused on introducing linking in general and C-V linking in particular. After discussing linking and its importance to speech perception and production, its mechanism was explained and students were asked to listen to and repeat several examples of C-V linking. Then, they were presented with instructions on how to answer the questions in the online training unit and worked on sample exercises. Students in the AV group received an additional 10 minutes of presentation that introduced waveforms, explained how they could reflect linking, and provided examples of linked and unlinked waveforms. Additionally, students had the opportunity to practice sample online questions that used waveforms as visual feedback. They were given a handout that included step-by-step directions on how to use Audacity to record their speech, analyze their waveforms, and compare them to those of the NS model (Appendix III for AV group training handout).

Similarly, day 3 and day 4 of the training started with shorter presentations that lasted between 10-15 minutes. The presentation on day 2 focused on V-V linking while that on day 3 focused on /h/ deletion. Each presentation started with a review of the previous session followed by a discussion of the specific linking aspect and examples. Students were then asked to work on sample exercises taken from the online unit. During practice time, the researcher went around the room to observe the students and answer any questions. In all presentations, the use of technical language in the training was minimal with more emphasis on practical issues.

## **Chapter Summary**

This chapter presented an overview of the research methodology employed in the dissertation study. It first explained how the quasi-experimental design was applied to investigate the effectiveness of the AO and AV training approaches on learners' linking perception and production. Following the description of participants and context, the data collection procedure and study timeline were presented. Next was a detailed account of the teaching materials developed for the AO and AV experimental groups that was accompanied by guidelines implemented for selecting and creating the training stimuli. The next three chapters detail the data collection tools and analysis procedure as well as the corresponding findings and discussion for linking perception, production, and learner perspectives of the two training approaches.

## **CHAPTER IV: THE EFFECTIVENESS OF LINKING INSTRUCTION ON PERCEPTION**

Each chapter of the next three chapters discusses one of the three major themes in the dissertation: (1) the effectiveness of instruction on linking perception, (2) the effectiveness of instruction on linking production, and (3) learners' perspectives of the effectiveness of the two proposed approaches to teaching linking. Chapter three detailed the methodology of the study including the research design, the characteristics of the participants, the data collection procedure and an account of the training materials. This chapter starts by describing how the perception pretest, posttest, and delayed posttest stimuli were developed in the study. It furthermore explains how participants' perception data were prepared for analysis, and describes the statistical measures that were utilized to answer every research question. Finally, it details the findings of linking perception tests and discusses them in the light of relevant literature.

### **Materials**

#### **Perception tests**

Participants took a perception test before and after the training sessions and one month later in order to evaluate their linking perception. The perception pretest, posttest, and delayed posttest included a dictation task where students listened to an audio recording and wrote sentences down on the provided worksheet. The audio was recorded by a native speaker of American English who was a graduate student in the TESL MA program. She was instructed to read the sentences twice as fluently and naturally as possible. The second reading was used for the perception tests. Sentences were numbered in the recording and on paper (see Appendix IV for perception test worksheet). There were six-to-nine second gaps between sentences to allow participants enough time to write them down. The length of the gaps was commensurate with the length of the sentences. All sentences were compiled in one audio file that was uploaded online. Using headphones, the participants listened to the sentences and wrote them down on the provided worksheet. The perception test audio recordings were available to students only during the time of the tests. Students were not able to control the audio or stop it at any time.

## Perception Test Stimuli

The pretest included 13 sentences that included 35 links. The first sentence in the pretest, for example, *Ames is a city in the west*, included two C-V linking pairs *Ames is* and *is a*, and one V-V linking pair *city in*. The posttest and the delayed posttest were identical and included the same sentences from the pretest in addition to 13 new unpracticed ones that had 29 additional links, a total of 26 sentences and 64 links (see Appendix V for perception tests stimuli). The sentences in the posttest were randomized and did not follow the same order of the ones in the pretest. The new sentences were utilized to explore how participants heard linked words in novel contexts not dealt with in the pretest or practiced in the training. The stimuli contained 75% C-V linking and 25% V-V linking (see Appendix VI for perception tests linking pairs).

Sentences were adapted from a number of ESL textbooks and real-life contexts. They were selected based on their word frequency and linking frequency. All words in the perception tests were from the K1 word list so that the participants were familiar with them. I piloted the stimuli as a part of two previous studies on linking perception. In those studies, I asked the NNS participants to indicate familiarity with the content words in the perception tests stimuli. Only sentences with words that 85% of the participants agreed on understanding were retained in the pool. In addition, care was taken so that every sentence included at least one linking pair. After piloting the perception test stimuli, some sentences were excluded because NSs did not naturally link some word pairs in them. The participants did not receive auditory or visual feedback during tests.

## Analysis

Participants' dictated sentences in the perception pretest, posttest, and delayed posttest were analyzed for accuracy in representing linking pairs. For a linking pair to be considered appropriately perceived, both words needed to be spelled correctly. If one word was missing or misspelled, the linking pair was considered wrong. All data were entered in an Excel sheet where correct pairs were represented by 1 and incorrect pairs were entered the same way students wrote them. The total and mean percentage score of correct links for every student were calculated for every test. In addition, I performed an item analysis to examine what made certain linking pairs

easier or more difficult to understand than others. The item analysis was conducted by calculating the mean percentage score of correct links for every linking pair on the posttest.

The independent variables of the study for the perception part were treatment group (CG, AO, AV), time (pretest, posttest, delayed posttest), and context novelty (old texts, novel texts); while the dependent variable was the mean percentage score of perceived linking pairs. Descriptive and inferential statistics were used to answer the three research questions. Table 4 lays out the research questions of the study, their variables, and the research method(s) used to answer every one of them.

Table 4. Perception research questions, variables and analysis methods

Research Question	Independent Variable	Dependent Variable	Analysis method
1. How effective are AO and AV training in improving linking perception for non-native learners of English?	Treatment (CG, AO, AV)  Time (pretest, posttest)	Change of scores from pretest to posttest (old items)	One-way ANOVA  Post-hoc: Scheffé test
2. How is the perception improvement caused by AO and AV training sustained over time?	Treatment (AO, AV)  Time (pretest, posttest, delayed posttest)	Scores of pretest vs. Scores of posttest (old items) vs. scores of delayed test (old items)	Two-way repeated measures ANOVA test 3x2  Post-hoc: one-way repeated measures ANOVA on each group
3. How do AO and AV training generalize to novel perception contexts?	Treatment (AO, AV)  Novelty (old items, new items)	Scores of posttest (old items) vs. Scores of posttest (new items)	Paired <i>t</i> -test on each group

To answer the first research question as to how AO feedback and AV feedback were effective in improving linking perception for NNSs, a one-way ANOVA was conducted to compare the change of scores between the pretest and posttest for the three groups. If the



difference was found to be significant, a post-hoc Scheffé test was used to further investigate the differences among groups.

To answer the second research question as to how perception improvement caused by training with AO feedback or AV feedback persisted, a two-way 3x2 repeated measures ANOVA test was used to examine the overall time effect on students' performance from the pretest to the delayed posttest. This test measured participants' scores on each of the tests and indicated whether their overall mean scores differed significantly. I looked at group-time interaction to examine the difference between the two experimental groups as well as the effect of time on the performance of each group. If differences were significant, a repeated-measures one-way ANOVA was employed for each group as a post-hoc test to further examine the differences.

To answer the third research question as to how improvement in linking perception transferred to novel contexts, AO and AV participants' scores on posttest old and new contexts were compared using paired *t*-tests. A bar graph and a table were used to report the results of these tests.

## Results

To examine the variability among groups before training, a one-way Analysis of Variance (ANOVA) was conducted on the participants' scores of the diagnostic listening test they had taken at the beginning of the semester (N=45). The ANOVA was not significant,  $F(2, 41) = 1.65, p = .20$ . Thus, there was a no significant difference in CG, AO, and AV group participants' starting listening proficiency before the training. The same outcome was found for participants' scores on the connected speech section in the same test,  $F(2, 41) = 1.65, p = .55$ . The results demonstrated that participants from the three groups had comparable proficiency levels in their overall listening skills as well as their connected speech skills before the onset of training. Any differences in participants' scores thereafter could be attributed to the treatment.

### Research question 1: The effectiveness of training

To answer the first research question as to how AO feedback and AV feedback were effective in improving linking perception for NNSs, the mean percentage scores of the CG, AO, and AV groups in the pretest and posttest were compared. The comparison involved their scores on the part of the pretest that was repeated in the posttest (old part). Table 5 demonstrates that all three groups (CG, AO, and AV) started with a similar mean percentage score in the perception pretest.

Table 5. Pretest and posttest mean scores of linking perception for the all groups

	<b>Control (N = 15)</b>		<b>Audio-only (N = 15)</b>		<b>Audio-visual (N = 15)</b>	
	Mean	SD	Mean	SD	Mean	SD
Pretest	64.38	16.76	62.28	14.89	63.23	16.77
Posttest	66.86	16.40	67.98	12.64	72.95	17.48

All groups obtained a greater score in the posttest increasing their percentage mean score by CG = 2.47, AO = 5.71, and AV = 9.71. However, the two experimental groups had a greater increase than did the control group after training. A comparison of groups' pretest and posttest mean scores can be seen in Figure 12.

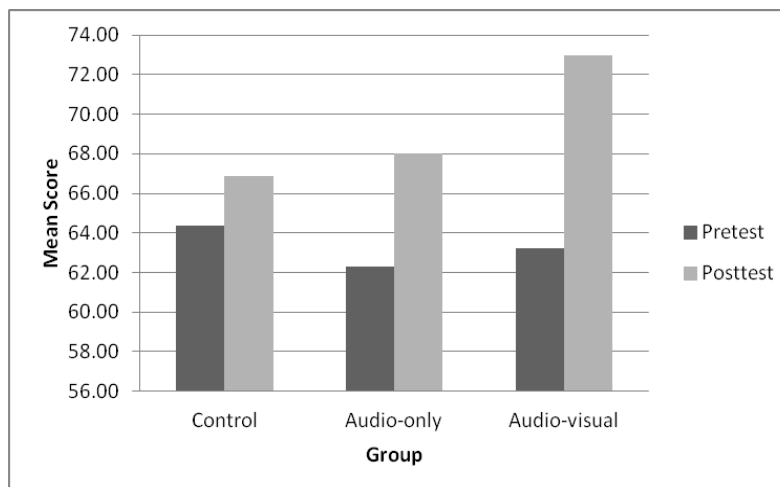


Figure 12. Mean perception scores for CG, AO and AV groups at pretest and posttest.

A one-way ANOVA was conducted on the change (or gain) of mean percentage scores between the pretest and posttest for the three groups with group as the independent variable. The assumption of the homogeneity of variance was tested and found tenable using Levene's Test,  $F(2, 42) = 2.74, p = .076$ . The ANOVA was not significant,  $F(2, 42) = 2.67, p = .081, \eta^2 = .11$ . Thus, there is no significant difference in CG, AO and AV participants' perception change scores ( $\alpha = .05$ ). There is a small difference among the groups based on Cohen's (1988) conventions for interpreting effect size. Since the overall ANOVA was not significant, no post hoc test was needed. To conclude, both AO and AV feedback improved learners' perception of linking in the dictation task more than the control group; however, the increase in scores was not significant. In addition, the AV group linking perception improved more than that of the AO group.

### **Research question 2: The long-term effects of training**

To answer the second research question as to how linking perception improvement caused by training with AO feedback or AV feedback persisted, a two-way 3x2 repeated measures ANOVA was used to examine the overall time effect on students' performance from the pretest to the delayed posttest. This test measures AO and AV participants' scores on each of the tests and indicates whether their overall means differ significantly. The results, displayed in Table 6, revealed no significant group-time interaction  $F(2, 56) = .11, p = .29$ .

Table 6. Results of two-way ANOVA test to examine long-term effect of treatment

	<b>Sum of squares</b>	<b>df</b>	<b>Mean square</b>	<b>F</b>	<b>Sig.</b>
Time	1267.98	2	633.99	1.25	.000
Group-time interaction	109.90	2	54.95	.11	.294
Error (time)	2455.99	56	43.85	-	-

The results indicate that the performance of the participants of two groups did not differ significantly over time. Figure 13 shows the performance pattern of the experimental groups from the pretest to the delayed posttest.

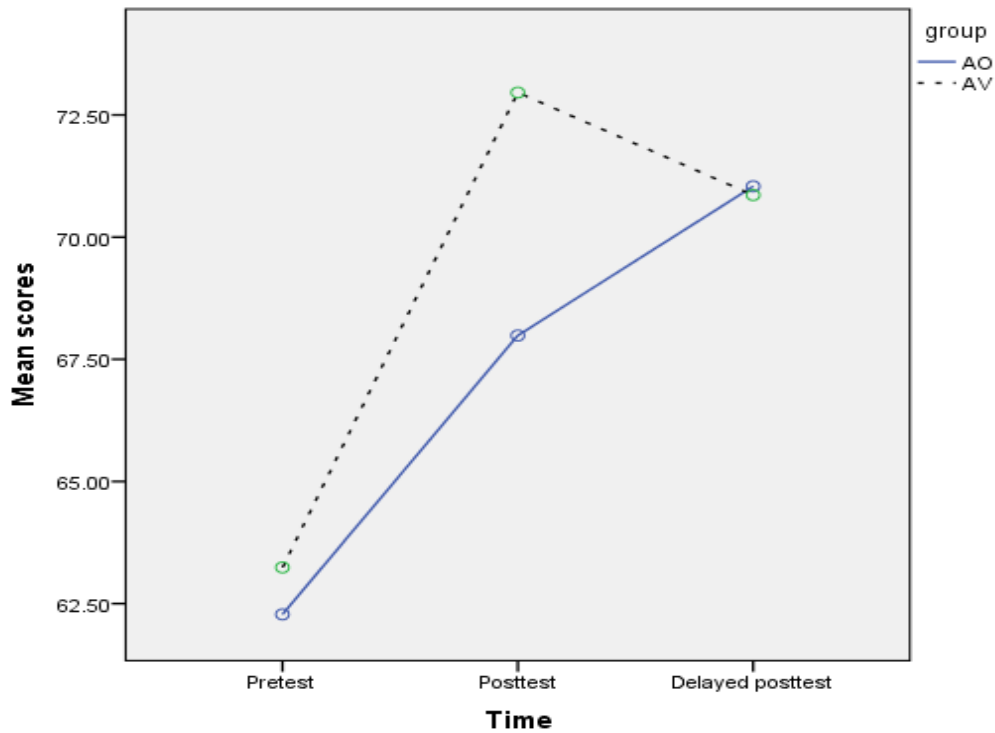


Figure 13. Perception scores for pretest, posttest, and delayed posttest for AO and AV groups.

However, as Table 6 showed, there was a ,  $F(2, 56) = 1.25, p < .001$ . This result indicates that the perception scores of one or both groups differed significantly between the pretest and delayed posttest. The test, nonetheless, does not reveal where exactly the difference lies.

To evaluate progress over time within each group, the students' mean scores on the pretest, posttest, and delayed posttest were compared for the AO group first, then the AV group. Table 7 provides the mean percentage scores and standard deviations for the three tests for both groups. In addition, it reports the results of the one-way ANOVA tests conducted on the two groups below.

In order to examine the overall time effect on the AO students' scores, a one-way repeated measures ANOVA was used to compare AO participants' performance on the three tests. The results indicated that learners' performance differed significantly over time,  $F(2, 28) = 8.91, p < .001, \eta^2_p = .39$ . Thus there is significant evidence that there was a change in participants' scores. This reveals that AO students' linking perception changed significantly over time from pretest to delayed posttest.

Table 7. AO and AV one-way repeated measures ANOVA results

	Test	Mean Score	SD	Sum of squares	df	Mean square	F	Sig.
<b>AO</b>	Pretest	62.28	14.89	593.13	2	296.56	8.91	.001
	Posttest	67.98	12.64					
	Delayed	71.04	11.78					
<b>AV</b>	Pretest	63.23	16.77	784.76	2	392.38	7.20	.003
	Posttest	72.95	17.48					
	Delayed	70.86	15.43					

To determine which tests were significantly different from each other, a pairwise comparison showed that there was a significant increase in AO scores over time from the pretest to the delayed posttest. The participants' mean percentage score was 62.28 at pretest, 67.98 at posttest, and 71.04 at delayed posttest. The pairwise comparisons revealed that students' change scores between pretest and posttest were not significantly different,  $p = .066$ , and between posttest and delayed posttest were also were not significantly different,  $p = .238$ . Nevertheless, students' change scores between pretest and delayed posttest were significantly different,  $p = .008$ . From the results, we conclude that students' proficiency in linking perception improved immediately after training with AO feedback, and continued to improve significantly by the time of the delayed posttest.

Similarly, in examining the overall time effect on the AV students' perception scores, the results of the one-way repeated measures ANOVA indicated that students' performance differed significantly over time,  $F(2, 28) = 7.20$ ,  $p < .003$ ,  $\eta^2_p = .34$ . This reveals that AV students' linking perception changed significantly over time from pretest to delayed posttest (Table 7).

The pairwise comparisons revealed that there was a significant increase in AV scores over time,  $p < .05$ . The participants' mean percentage score was 63.23 at pretest, 72.95 at posttest and 70.86 at delayed posttest. The pairwise comparisons showed that students' change scores

between pretest and posttest were significantly different,  $p < .008$ . Yet, students' change scores between pretest and delayed posttest were not significantly different,  $p = .059$ , and the same applied to the difference between the scores of the posttest and delayed posttest,  $p = 1.00$ . Students' proficiency in linking perception significantly improved immediately after training with AV feedback, and then slightly declined by the time of the delayed posttest.

To conclude, the results indicate AO and AV training were effective in improving students' linking perception beyond the duration of training. After improving their scores from the pretest to the posttest, both AO and AV students were able to retain the majority of this improvement when evaluated one month after the end of training. While the performance of the AO group continued to improve after the posttest, AV group performance declined slightly by the time of the delayed posttest one month after the end of training.

### **Research question 3: Transfer to novel contexts**

To answer the third research question as to how improvement in linking perception transferred to novel contexts, I compared students' mean scores on the part of the posttest that had the old items presented in the pretest to their score on the part of the posttest that had the new items added to the posttest. A paired sample  $t$ -test was conducted to evaluate whether a statistically significant difference existed between the mean scores on the old and new parts of the posttest for the AO group then the AV group. Table 8 presents the differences between the two tests, the difference values, the  $t$ -values and their degrees of freedom ( $df$ ), and the significance of the tests ( $p$ -value) for both groups.

The results of the paired sample  $t$ -test were not significant for the AO group,  $t(14) = .93$ ,  $p = .36$ ,  $\eta^2 = .17$ , indicating that there was no significant difference between the posttest old score and the posttest new score. The perception improvement that the AO students achieved on the new part of the posttest decreased from  $M = 67.98$  to  $M = 65.06$ . The mean decrease was 2.92.

Table 8. Results of paired *t*-tests on perception posttests for old and novel contexts

	Context Novelty	Mean score	SD	Mean difference	<i>df</i>	<i>t</i>	Sig.
<b>AO</b>	Old	67.98	12.64	2.92	14	0.93	.364
	New	65.06	20.28				
<b>AV</b>	Old	72.96	17.48	.30	14	.11	.912
	New	72.65	17.46				

Furthermore, the results of the paired sample *t*-test were not significant for the AV group,  $t(14) = .11$ ,  $p = .91$ ,  $\eta^2 = .02$ , showing that there was not a significant difference between the posttest old score and the posttest new score. The perception improvement that the AO students achieved in the repeated part of the posttest decreased very slightly from  $M = 72.96$  to  $M = 72.56$  in the new test with no significant change. The mean decrease was only .30 percent. Figure 14 shows the differences in the performance of the two experimental groups on old and novel items.

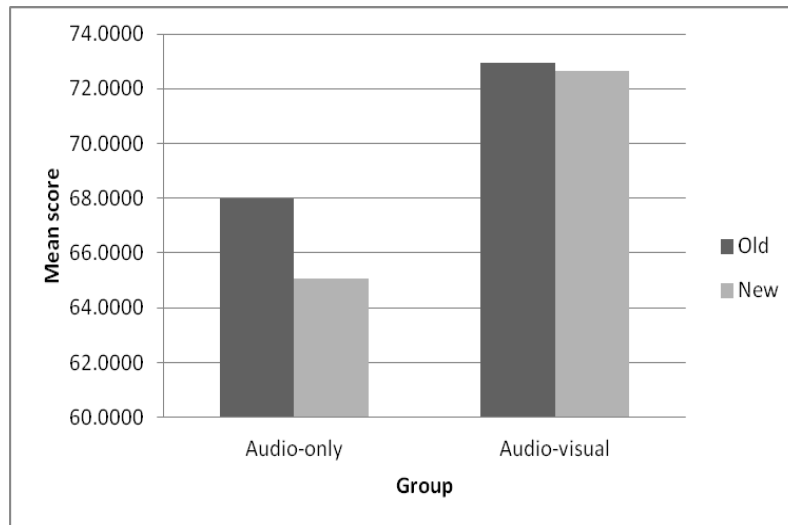


Figure 14. Perception scores for posttest old and novel items for AO and AV groups

The results of no significant difference between scores of old and new items translate into significant retention of improvement. In general, AO and AV participants improved their linking on the new items slightly less than they did with the old items in the perception posttest.

However, the decrease in the new-items scores was not significant. On the other hand, AV training helped to transfer the gained improvement to novel contexts better than AO training.

## **Discussion**

This section offers a summary and discussion of the major findings in the perception part of the dissertation. It presents an analysis of the results in the light of relevant previous research. In addition, the section closely examines examples from the data to provide an in-depth understanding of students' perception of linking. The section is organized by research questions.

### **Research question 1: The effectiveness of training**

The first research question investigated how AO and AV training were effective in improving linking perception for non-native learners of English. To determine the effectiveness of each training approach, 45 students took a dictation task where they listened to the same sentences before and after training (sentences were randomized in posttest). The students participated in a two-week training that involved four 50-minute sessions during their regular classes. The training included in-class and online materials with activities to improve their perception and production of English linking. Throughout the online training, students in the two experimental groups received either audio-only or audio-visual feedback on their performance depending on the group they belonged to.

Results of a one-way ANOVA indicated that there was no significant difference in CG, AO, and AV participants' perception scores in the pretest and posttest. All groups obtained a greater score in the posttest increasing their percentage mean score by CG = 2.47, AO = 5.71, and AV = 9.71. AO and AV feedback improved learners' perception of linking more than the control group, yet the improvement was not significant. Furthermore, AV training helped improve participants' linking perception better than AO training.

The AV group outperformed the AO group in linking perception. This is in line with previous research which concluded that AV training involving EVF improved speech perception and production greater than AO training. It also agrees with previous research that found that learners who received both audio and visual feedback performed significantly better than



learners who received audio feedback only (de Bot, 1983; Hardison, 2004; Motohashi-Saigo & Hardison, 2009).

The findings are also relatively consistent with previous studies on CSP perception that showed positive effect of instruction on the perceptual performance of NNSs. Following is a summary of three research studies that investigated the effectiveness of language learning materials on CSP perceptual accuracy (Brown & Hilferty, 1986b; Crawford, 2006; Kuo, 2009). These studies were selected out of eight outlined in the literature review because of their relevance to the study, well-reported methods, and/or a control-group design.

Brown and Hilferty (1986) recruited 32 Chinese L1 graduate students who were studying in the US. The participants were randomly assigned to a control group and an experimental group. The experimental group received four weeks of daily ten-minute lessons that involved exposing the students to the new forms, and then having them check their comprehension of the forms with dictation exercises. The results showed that, in the posttest, participants in the experimental group increased their ability to identify and write down reductions by 32% over the pretest.

Building on Brown and Hilferty's study, Crawford (2006) mainly examined whether weekly perceptual training sessions with reduced forms led to improvement of learners' perceptual accuracy of these forms. With 26 Japanese students in the control group and 23 in the experimental group, the study implemented a pretest/posttest research design. The two tests used the same 20-item cloze dictation test. The training focused on teaching specific high frequency reductions undergoing various CSPs, for example *wanna* (want to), *doncha* (don't you), *'e* (he), and *couldja* (could you), and followed the same format of the pretest and posttest. After seven 15-minute training sessions spread out on seven weeks, the experimental group showed significantly higher gains than the control group where most of the reductions improved by 30 percent.

More recently, Kuo (2009) investigated the effect of English linking instruction on elementary Taiwanese school students' speech production and phonological awareness. The CG (N = 32) received no treatment, while the experimental group (N = 33) received 40-minute training sessions on C-V, V-V and C-C linking. The pretest and posttest had four sections: (1)

“Marking linking” to investigate whether the participants were familiarized with linking characteristics, (2) “Listen and Fill in the Blanks” to evaluate students’ ability to interpret the stream of connected speech, (3) “Sentence Reading Test” to evaluate participants’ ability to read sentences with linking and measure their phonological awareness on phoneme blending by making sound change, and (4) “Listen and Circle” to test the students’ listening and phonological awareness in terms of students’ understanding of the sound structure. After 14 weeks of training, results indicated that the experimental group significantly improved their speech production and developed phonological awareness. The significant improvements were made in all subtests except for “Listen and Circle.”

In the present study, students in the experimental groups improved their linking perception after training which suggested that AO and AV training were effective in increasing students’ awareness of the phenomenon of linking and rate of noticing linked words. The improvement, however, was not statistically significant which can be attributed to several reasons.

First, exact comparisons between the previous studies and the present study are somewhat challenging because there are a number of differences, most importantly, in the length and distribution of training and the target CSP(s). It may be that the longer training time in Kuo (2009) and the more balanced distribution of training sessions in Brown and Hilferty (1986) and Crawford (2006) contributed to higher perceptual gains. The present study was held over four 40-minute sessions over the period of two weeks. The period of training was limited to accommodate the other materials in the course syllabus. It was not possible to spread the training over the whole semester due to schedule conflicts. Lack of adequate training time may have contributed to insignificant gains in students’ perception scores.

Comparisons with Brown and Hilferty’s study are also challenging to make because the authors did not specifically indicate the types of reductions their students were trained on, although they were likely to be similar to the ones in Crawford: a selected set of high frequency reductions, such as *gonna*. The perceptual competence needed to recognize C-V and V-V linking can be different and more challenging to acquire than that needed to recognize a predefined set of reductions. It is easier for the listener to look for a formulaic expression stored in the memory, such as *wouldja*, than to compute it (Ellis, Simpson-Vlach, & Maynard, 2008). Both NSs and

NNSs have available to them “a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments” (Sinclair, 1991, p. 100) which facilitates the recognition of such high-frequency expressions. However, because linking is a pervasive CSP that can occur at any C-V or V-V word boundaries, recognizing it in speech requires an additional set of skills. This makes learning linking more challenging than other pre-structured reductions. A rule-based process, in general, reduces the automaticity in perception.

A closer examination of the data provides further insights about the effect of the nature of linked words on their linking potential and learnability. Table 9 includes percentage mean scores of selected linked words from the pretest and posttest and the difference between the two tests for AO and AV groups and their combined scores (Appendix IX includes the same details about all linked words).

Table 9. Item analysis of perception mean percentage scores

Items	AO			AV			Combined		
	Pretest	Posttest	Difference	Pretest	Posttest	Difference	Pretest	Posttest	Difference
puts on	93	93	0	93	93	0	93	93	0
come over	93	80	-13	87	100	13	90	90	0
come up	93	100	7	87	100	13	90	100	10
with an	33	40	7	27	47	20	30	43	13
an answer	33	40	7	27	47	20	30	43	13
too often	80	93	13	93	93	0	87	93	7
see her	0	0	0	7	7	0	3	3	0
her in	7	0	-7	40	7	-33	23	3	-20
in an	80	93	13	67	93	27	73	93	20
an easy	67	73	6	53	73	20	60	73	13
easy example	100	100	0	100	100	0	100	100	0
looked at	20	20	0	33	33	0	27	27	0

The pretest scores suggest that some words were easier to link than others, e.g. *puts on*, *come over*, *come up*, *too often*, and *easy example*. Because students' pretest scores were high for these items, there was little room for improvement after training. Most of these words are high frequency chunks that are usually recognizable as a unit that listeners store as a single semantic and phonetic entity (Field, 2003b). In the case of NNSs, some chunks are more recognizable than others depending on their frequency. This can be demonstrated in the case of *come up with* where students scored high on the first part of the phrasal verb *come up* (combined M = 90), but low on the second part *with an* (combined M = 30). It seems that most students did not anticipate hearing *with* after *come up* so they failed to write it. This can be attributed to the difference in the frequency of *come up* and *come up with*. One exception was *looked at* which proved to be challenging to students even after training (combined M = 27, no improvement). For this sentence, most students wrote *he looked/look everything around him* instead of *he looked at everything around him*. This contrasts with the aforementioned frequency-based explanation because *look at* is even more frequent than *come over* and *come up*. In this case, the past tense morpheme *-ed* could be the source of confusion since it was pronounced as /t/ which, in turn, was confused with *at*.

The item analysis of perception scores also revealed that the mean scores of most students were low in the pretest for *with an*, *an answer*, *see her*, and *her in*. Instead of *with an answer*, many students wrote *with a answer*. Similar to *looked at*, the phonological environment might have influenced students' decision who confused the indefinite article *an* with the first syllable in *answer*. However, it seems that the training had positive effects on the perception of items containing *an*, such as *in an* and *with an* since their mean scores increased by 13 percent and 20 percent respectively. This finding suggests that the perception of linked *an* was more learnable than other words in the study.

An important factor that contributed to the difficulty level of linking perception is the simultaneous presence of other CSPs, such as vowel reduction, flapping and deletion. Listeners managed to understand such combinations when they existed in high frequency chunks, such as *get out* that involved linking and flapping. However L2 learners were challenged when they came across less frequent combinations, such as *with an* and *see her*. The latter was especially challenging because of the drastic change that took place within the two words due to /h/

deletion, V-V linking, insertion of the glide /j/, and light /r/. Indeed, previous research indicated that /h/ deletion and V-V linking were difficult to learn compared to other types of CSPs (Crawford, 2006; Kuo, 2009) due to the interaction of different variables.

Previous studies on CPS perception, coupled with the findings of the present study, show the positive effect of instruction on L2 learners' perception of linking. The study also demonstrated that training with the help of audio-visual feedback resulted in higher perception rates than audio-only training. However, the results were not statistically significant which may be attributed to lack of adequate training time and potential factors affecting the ease and learnability of linking perception. Such factors include frequency of linked words, their phonological environment and the presence of other CSPs at the same link.

### **Research question 2: The long-term effects of training**

The second research question investigated how linking perception improvement caused by training with AO feedback or AV feedback persisted. To evaluate training long-term effects, students took a delayed posttest one month after the end of the training. The mean scores of the two experimental groups were compared using a two-way repeated measures ANOVA that was followed by post hoc pairwise comparisons to pinpoint significant differences. Results revealed that there was no significant difference between the scores of the two groups over time. However, both AO and AV training were effective in improving students' linking perception beyond the duration of training. While the performance of the AO group continued to improve after the posttest by 3.04 percent, the AV group performance declined slightly by 2.09 percent one month after the end of training yet remained above their pretest mean score.

Results of the previous studies generally indicated that CSP instruction facilitated learners' perception of connected speech. However, these studies failed to address the long-term effects of such training on learners' perceptual accuracy. It is, therefore, challenging to make comparisons. But by observing the general trend of delayed test scores in L2 pronunciation studies, it becomes apparent that they were in line with the results of the AV group. L2 students who are able to improve certain aspects in their pronunciation after training generally undergo phonological "backsliding" (Beebe, 1988; Sardegna, 2011) or decrease in their training gains. As long as they retain an improvement level well above the pre-training level, it can be said that the

training was effective in improving the target pronunciation skill(s) over time. This was true in the case of AV students whose performance dropped slightly after one month of the training.

The result of the AO group is somewhat surprising because after a relatively small improvement in the posttest, the students' performance continued to improve in the delayed posttest and was significantly higher than their starting point. A number of reasons could have contributed to this unexpected result. Due to the short period of training, learners may have not been able to take full advantage of what they had learned. Although no more linking practice was offered between the posttest and the delayed posttest, students may have independently incorporated the newly acquired linking knowledge in their ESL classes and daily activities. Thus, they were able to improve their performance of linking perception. Another possible reason for such an improvement may be fluctuating motivation level. Informal class observations showed that several students from the AO group were not motivated to work on the training activities, likely because they were not graded and did not count toward their final course grade. It is possible that students put more effort in their class work in general toward the end of the semester. Finally such inconsistency in performance may be attributed to the small sample size of the participants in the study.

### **Research question 3: Transfer to novel contexts**

The third research question examined how improvement in linking perception transferred to novel contexts that were not on the pretest. As demonstrated in the previous research questions, participants in the study improved their linking perception after training. However, to evaluate the generalization of training improvement to new words, the perception posttest had another part with sentences not included in the pretest. Such a test also validates any test effect resulting from repeated use of the original pretest over time. If students' scores on both the old and the new parts of the posttest were significantly different, then it is likely that their prior knowledge of the test items affected their listening perception. Such a result could make their performance on the old posttest and delayed posttest not reliable and affect the generalizability of the findings.

The mean scores of the old and new parts of the posttest were compared using paired sample *t*-tests for the AO and AV groups. Results were not significant for both groups with a

mean score decrease of 2.92 for the AO group and .30 for the AV group. This slight decrease in scores translates into significant transfer of improvement to new contexts. Since the difference was not significant, it is likely that the performance of the groups on both tests reflected their actual knowledge of linking perception skills and was not the result of any test effect.

The overall perception scores in novel contexts indicated that AV training was more effective than AO training in improving learners' perception of linking in novel contexts. The findings are in line with previous research that supported the ability of learners to generalize improved perceptual capabilities to novel stimuli (e.g., Hardison, 2003; Hirata, 2004; Lively, Logan & Pisoni, 1993; Motohashi-Saigo & Hardison, 2009; Ruellot, 2006). Such positive transfer of improvement is the goal of any pedagogical L2 research.

## **Chapter Summary**

This chapter examined the effectiveness of instruction on linking perception using AO and AV feedback. It started by describing the perception instrument of the study including the pretest, posttest, and delayed posttest stimuli. It then explained how participants' perception data were prepared for analysis, and described the statistical analysis scheme. Finally, the findings of this part of the study were discussed in the light of relevant literature.

The results revealed a positive effect of instruction on L2 learners' perception of linking for both groups immediately after and one month after the end of training. The training was also successful in helping students transfer gained improvement to novel contexts. Although this finding suggests that students can generalize perceptual gains, such finding should be considered in relation to the small sample size of the study and may need further research to investigate the results further.

The chapter also analyzed specific examples from the study and discussed factors that could affect the ease and learnability of linking perception, such as frequency of linked words, their phonological environment and the presence of other CSPs at the same link. Finally, the results of this study provide support for the inclusion of more effective approaches, such as audio-visual feedback, to the teaching of linking perception rather than continuing to exclusively rely on more traditional listening teaching methods.

## **CHAPTER V: THE EFFECTIVENESS OF LINKING INSTRUCTION ON PRODUCTION**

This chapter discusses the research materials and analysis employed to investigate the effectiveness of instruction on the production of linking. Chapter three detailed the methodology of the study including the research design, the characteristics of the participants, the data collection procedure, and an account of the training materials. This chapter starts by describing the test stimuli of the production pretest, posttest, and delayed posttest that were used to collect the production data of the study. This is followed by an explanation of how participants' production data were prepared for analysis and then analyzed. It furthermore describes the statistical procedure that was utilized to answer the four research questions. Finally, the chapter presents the results and discusses their significance in the light of the context of the study and relevant literature on connected speech production.

### **Materials**

#### **Production tests**

Three production tests were administered to students before, immediately after, and one month after training in order to evaluate their linking proficiency and the effectiveness of AO and AV training on their performance. A text reading task was chosen for the tests because it allowed for controlling linking frequency and types, pausing, vocabulary, and grammar. The task was divided into two parts: the first part was read in the pretest, and the second part was read in the pretest, posttest, and delayed posttest (see Appendix VII for production test stimuli). The purpose of second (new) part was to examine whether there was a transfer of linking skills to novel contexts that were not included in the pretest. The reading text stimuli involved a balanced distribution of C-V and V-V linking (see Table 11).

In order to examine the impact of word frequency on linking, two texts were used in each part of the production test. The first text was a fiction story about a boy and a snake that was mainly comprised of K1 high frequency (HF) words. The story was originally written by a NS and modified by the researcher to include only K1 words. The second text was a short essay about language acquisition that included a high percentage of low frequency (LF) words. The



purpose of using the short essay was to investigate whether LF words were more difficult to link to and from than HF words. The essay was written specifically as a stimulus for the production test. After deciding on an outline for the essay, a number of LF vowel-initial words were selected from the Academic Word List (Coxhead, 2000) to be incorporated in the text. This would create opportunities for linking to and from vowel-initial LF words.

The two texts appeared on two separate pages, taking into account that all linked words appeared on the same line. This was to avoid having students break the link at the end of the line due to the time it took them to move their eyes to the beginning of the following line. Table 10 highlights differences between the HF and LF texts in terms of C-V and V-V linking frequency and pausing frequency.

Table 10. Potential linking and phrasing in production stimuli

	Linking pairs (Per 100 words)	C- V (Per 100 words)	V-V (Per 100 words)	Pauses (Per 100 words)
HF text (complete)	25.5	19.6	6	12
Pretest HF text	28	22	6	13
Novel HF text	22.6	16.7	6	10.7
LF text (complete)	31.2	23.2	8	11.6
Pretest LF text	31	25.4	5.6	11.3
Novel LF text	31.3	20.9	10.4	11.9

### **Production tests stimuli**

Since NSs link most but not all their linking pairs (Alameen, 2007b), an analysis was needed to decide which linking pairs should be accounted for in the analysis of NNSs' recordings. In order to establish a baseline of required, optional, and unlinked contexts of linking, five NSs of American English were asked to provide speech samples of the reading texts. All five participants were females who had lived in Iowa for more than four years. Two of them had a bachelor's degree and three had a master's degree. Their ages ranged between 29 and 50. The NSs' reading was used in the present study for two purposes: 1) to mark the links that NSs actualized and, 2) to mark NS pauses on the text.

Only those links that two or more NSs agreed on linking were retained in the analysis pool. Other links were considered ‘unrequired’ and were hence not counted as links in the study (see Appendix VIII for a list of linking pairs in the production test stimuli). As Table 10 showed, the HF text included 46 linking pairs (25.5 links per 100 words), and the LF text included 43 (31.2 links per 100 words).

Since pausing prevents linking, the production test stimuli were controlled for pausing. For this purpose, red slashes were used to mark required pauses in the texts (see Appendix VII for pausing placement in the texts). The NNSs in the study were asked to pause at those slashes; however, pausing was not enforced on participants if they did not do it. In order to establish the places for required pauses, the production of the five NSs was analyzed for phrasing and pause placement. The required pauses were those that three out of five NSs agreed on. Table 11 displayed pausing frequency in the production test stimuli as well as linking frequency.

In order to control for vocabulary frequency in the production stimuli, the original texts were modified so as to contain a certain percentage of HF and LF words. Cobb’s Web Vocabulary Profiler v.4 (Cobb, n.d.; Heatley, Nation, & Coxhead, 2002) was utilized to break the texts down by word frequencies in the language at large. Furthermore, Flesch-Kincaid readability statistics were calculated for each text. Flesch Reading Ease test rates text on a 100-point scale. The higher the score, the easier it is to understand the document. Flesch-Kincaid grade level test rates text on a US school grade level. Most words in the HF text were from the K1 word level to ensure NNSs’ familiarity with all the words in the text. Furthermore, the text was deemed appropriate for this level of NNSs based on the results of a previous study, in which I piloted the stimuli on the participants. All of the students agreed that they understood all the words in the HF text. On the other hand, the LF text included a high percentage of LF words (18%) taken from the Academic Word List. Readability statistics can be seen in Table 11.

Table 11. Production test stimuli readability

	Number of words	K1 words %	K2 words	AWL words	Flesch Reading Ease	Flesch-Kincaid Grade Level
<b>HF text</b>	184	92%	8%	0%	100	2.1
<b>LF text</b>	138	77%	4%	18%	61.1	7.7

## **Analysis**

This section describes the procedures used to analyze the production data, the audio files of the study. This includes defining the analysis criteria used by the researcher and the second rater, and outlining statistical measures implemented to answer every research question.

### **Preparing production data**

The raw audio files from all tests were named in the following manner: ‘speaker’s name + test time + group’. Since students were instructed to say their names at the beginning of every recording, names were extracted from the files in addition to any other identifiers, such as the voice of the interviewer to guarantee anonymity. Every audio file was then divided into smaller files to allow for blind rating. The categories were:

- (1) HF passage: included the first part of the HF text,
- (2) HF passage new: included the second part of the HF text that was introduced in the posttest,
- (3) LF passage: included the first part of the LF text, and
- (4) LF passage new: included the second part of the LF text that was introduced in the posttest.

Every category included audio files from the pretest, , and delayed posttest. Every participant produced 10 audio files: pretest HF, pretest LF, posttest HF, posttest HF new, posttest LF, posttest LF new, delayed posttest HF, delayed posttest HF new, delayed posttest LF, delayed posttest LF new. CG participants had only six files each because they did not take the delayed posttest. The audio files within each of the four categories (HF, LF, HF new, LF new) were randomly coded and numbered so that the raters were not able to identify the group or time of the test. The original file names and their matching coded names were kept in a spreadsheet and were not provided to raters.

### **Rating criteria: Revisiting the definition of linking**

In rating participants’ linking pairs as linked or unlinked, I implemented the criteria in chapter one of the study where linking is (1) “connecting the boundaries of two words while

keeping their phonetic qualities,”( 2) “without detecting a glottal stop between the two words.” For example, if a student linked the two words *top of* by connecting their boundary sounds /p/ and /ə/, but inserted a glottal stop at the onset of /ə/, the two words were considered unlinked.

Nonetheless, applying these criteria was not a straightforward process. The part that states “while keeping their phonetic qualities” needed further elaboration. In many instances learners changed the pronunciation of boundary sounds of linked words. The extent to which such changes can be tolerated is debatable and contingent, to a great extent, upon how NSs modify boundary sounds in C-V and V-V linking. For example, when ‘*carried it*’ was linked as ‘*carry it*’, the two words were not counted as linked even though the learner connected the two words *carry* and *it* because the boundary sound /d/ was missing.

In other instances, one or both boundary sounds were modified either by prolonging a vowel (e.g., a longer /ɪ/ in *take it*, a reduced vowel or glide (e.g., a reduced /eɪ/ in *be able*), or a modified stress pattern (e.g., a stressed *it* in *take it*). Such modifications were deemed as minor and did not affect the rating of linking since I noticed that NSs performed similar processes when they linked similar phonological contexts. This variation of degree is a characteristic of other CSPs where speakers do not always produce a specific CSP in the same way (Barry, 1984, 1985). More research is needed to identify subtle characteristics of linking which will facilitate rating linking, in particular, and other CSPs, in general.

## **Second rater training**

After analyzing all the production audio files, 10% of the data were randomly selected to be rated by a second rater in order to check consistency in applying the rating criteria. The second rater was a NS graduate TESL professional who did a rater training for about 15 minutes. The training included an introduction to the concept of linking and its rating criteria. The criterion of “not detecting a glottal stop between the two words for linking to happen” was especially emphasized because in previous second ratings, the rater overlooked this criterion and relied on the conventional definition of linking. In addition, the rater was made aware of potentially confusing connected speech contexts, such as deleted /h/. The training involved an explanation of the task as well as audio examples. The rater was provided with a transcript of the recordings and an Excel sheet that contained the potential linking pairs. Inter-rater reliability was

calculated and used to examine the consistency and reliability of the human ratings across raters for NNSs' production. Intraclass correlation was used to calculate the inter-rater consistency and was found to be 0.82 which indicated a strong level of agreement between the ratings of the researcher and the second rater.

### **Production analysis procedure**

The data were analyzed with the help of Audacity, which removed unwanted noise and assisted in isolating linking pairs. Data were entered into an Excel sheet where linked pairs were represented by 1, and unlinked ones by 0. The total and mean score of linked pairs were calculated for every audio file. After that, files were matched with the participants' names and sorted into their original groups and test times. In addition, an item analysis was performed to examine and understand what makes certain linking pairs easier or more difficult to link than others. The item analysis was conducted by calculating the mean percentage score of correct links for every linking pair on the posttest.

For the production part, the independent variables were treatment group (CG, AO, AV), time (pretest, posttest, delayed posttest), novelty (old texts, new texts), and word frequency (HF words, LF words); while the dependent variable is the mean percentage score of actualized linking pairs. Descriptive and inferential statistics were used to answer the three research questions. Table 12 outlines the study's research questions, variables, and analysis methods for the production data.

To answer the first research question as to how AO feedback and AV feedback were effective in improving NNSs' linking production, a one-way ANOVA was conducted to compare the change of scores between the pretest and posttest for the control group and the two experimental groups. If the difference was significant, a post-hoc Scheffé test was used to further investigate the differences among groups. A bar graph and a table are used to report the results of these tests.

Table 12. Production research questions, variables and analysis methods

Research Question	Independent Variable	Dependent Variable	Analysis method
1. How effective are AO and AV training in improving linking production for non-native learners of English?	Treatment (CG, AO, AV) Time (pretest, posttest)	Change of scores from pretest to posttest (old items)	One-way ANOVA Post-hoc: Scheffé test
2. How is the production improvement caused by AO and AV training sustained over time?	Treatment (AO, AV) Time (pretest, posttest, delayed posttest)	Scores of pretest vs. Scores of posttest (old items) vs. scores of delayed test (old items)	Two-way repeated measures ANOVA test 3x2 Post-hoc: one-way repeated measures ANOVA on each group
3. How do AO and AV training generalize to novel production contexts?	Treatment (AO, AV) Novelty (old text, new text)	Scores of posttest (old items) vs. Scores of posttest (new items)	Paired <i>t</i> -test on each group
4a. Are low frequency words more difficult to link to and from than high frequency words?	Word frequency (HF text, LF text)	Scores of posttest old HF vs. Scores of posttest old LF	Paired <i>t</i> -test
4b. Do learners improve better after treatment in HF contexts than in LF contexts?		Change of scores from pretest to posttest (old HF) vs. Change of scores from pretest to posttest (old LF)	Paired <i>t</i> -test

To answer the second research question as to how production improvement caused by training with AO feedback or AV feedback persisted, a two-way 3x2 repeated measures

ANOVA test was used to examine the overall time effect on students' performance from the pretest to the delayed posttest. This test measured participants' scores on each of the three tests and indicates whether their overall means differed significantly. Group-time interaction was used to examine the difference between the two experimental groups as well as the effect of time on the performance of each group. If differences were significant, a repeated-measures one-way ANOVA and pairwise comparisons was devised for each group as a post-hoc test to further examine the differences. A plot and a table are used to report the results of these tests.

To answer the third research question as to how improvement in linking production transferred to novel contexts, AO and AV participants' scores on the old and new parts of the posttest were compared using two paired *t*-tests. A bar graph and a table are used to report the results of these tests.

To answer the first part of the fourth research question as to how word frequency affected linking production, a paired *t*-test was used to compare the mean percentage scores of linking to and from HF and LF words in the posttest for all participants. For the second part of the question that investigated whether learners' production of linking improved better in HF contexts than in LF ones, a paired *t*-test was conducted on the change of scores from pretest to posttest of the HF and LF texts in order to compare students' performance over time in HF and LF contexts.

## Results

### Research question 1: The effectiveness of training

To answer the first research question as to how effective AO feedback and AV feedback were in improving linking production for NNSs, the mean percentage scores of the CG, AO, and AV groups were compared in the pretest and posttest. The comparison involved their scores on the first part of the HF text that was repeated in the posttest. Table 13 displays the mean percentage scores and standard deviations of the CG, AO, and AV groups in the pretest and posttest.

Table 13. Pretest and posttest scores of linking production for all groups

	Control (N = 12)		Audio-only (N = 15)		Audio-visual (N = 15)	
	Mean	SD	Mean	SD	Mean	SD
Pretest	33.33	19.67	35.55	19.09	32.09	20.75
Posttest	37.03	18.54	55.55	22.76	50.86	13.49

Table 13 shows that all three groups (CG, AO, and AV) started with a similar mean percentage score in the production pretest. A one-way ANOVA run on the pretest scores revealed no significant difference among the three groups ( $p = .87$ ) which indicates that participants had comparable proficiency levels in their linking production skills before the beginning of training. Any differences in participants' scores thereafter can be attributed to the treatment.

At the time of the posttest, all groups obtained a greater score increasing their percentage mean scores by CG = 3.70, AO = 20, and AV = 18.76. However, the two experimental groups had a greater increase than that of the control group after training. A comparison of groups' pretest and posttest scores can be seen in Figure 15.

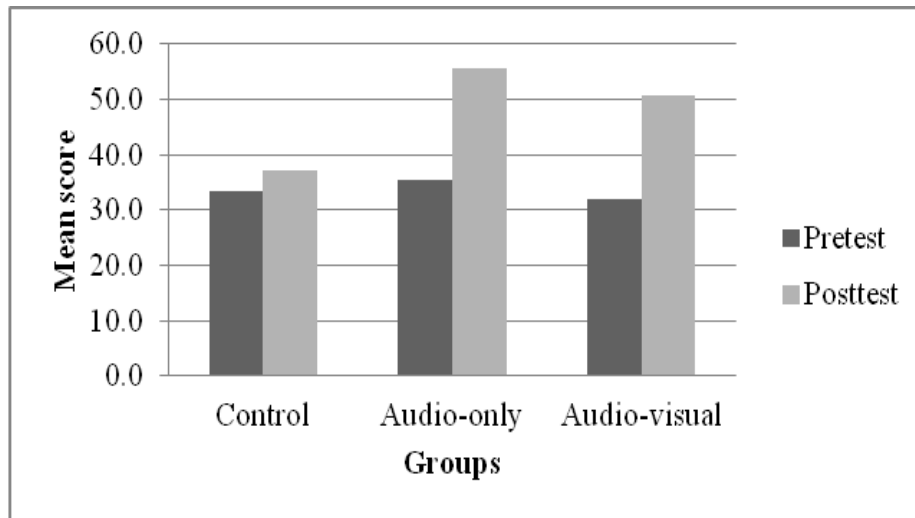


Figure 15. Mean production scores for CG, AO, and AV groups for pretest and posttest



A one-way ANOVA was conducted on the change of mean percentage scores between the pretest and posttest for the three groups with group as the independent variable. The assumption of the homogeneity of variance was tested and found tenable using Levene's Test,  $F(2, 39) = 2.64, p = .08$ . The ANOVA was significant,  $F(2, 39) = 6.84, p = .003, \eta^2 = .25$ . Thus, there was a significant difference in participants' production change ( $\alpha = .05$ ). There was a medium difference between the group based on Cohen's (1988) conventions for interpreting effect size.

Since the overall ANOVA was significant, multiple comparisons were implemented to locate significant differences in group performances. Post hoc comparisons to evaluate pairwise differences among group means were conducted with the use of Scheffé test since equal variances were tenable. Tests revealed significant pairwise differences between the mean scores of the participants from the CG and AO group ( $p = .007$ ) and between CG and AV group ( $p = .013$ ). The scores of the participants in the AO group did not significantly differ from those in the AV group, ( $p = .96$ ). To conclude, the increase in production scores of both AO and AV groups was significantly greater than that of the control group. The AO group linking production improved slightly more than that of the AV group.

### **Research question 2: The long-term effects of training**

To answer the second research question as to how linking production improvement caused by training with AO feedback or AV feedback persisted, a two-way 3x2 repeated measures ANOVA was used to examine the overall time effect on students' performance from the pretest to the delayed posttest. This test measures AO and AV participants' scores on each of the tests and indicates whether their overall means differ significantly. The results, displayed in Table 14, revealed no significant group-time interaction  $F(2, 56) = .11, p = .89$ .

Table 14. Results of two-way ANOVA test to examine long-term effect of treatment

	Sum of squares	df	Mean square	F	Sig.
Group-time interaction	18.45	2	9.23	.11	.896
Time	5903.60	2	2951.80	35.03	.000
Error (time)	4718.22	56	84.25	-	-

The results indicate that the performance of the participants of the two groups did not differ significantly over time. Figure 16 demonstrates the performance pattern of the experimental groups from the pretest to the delayed posttest.

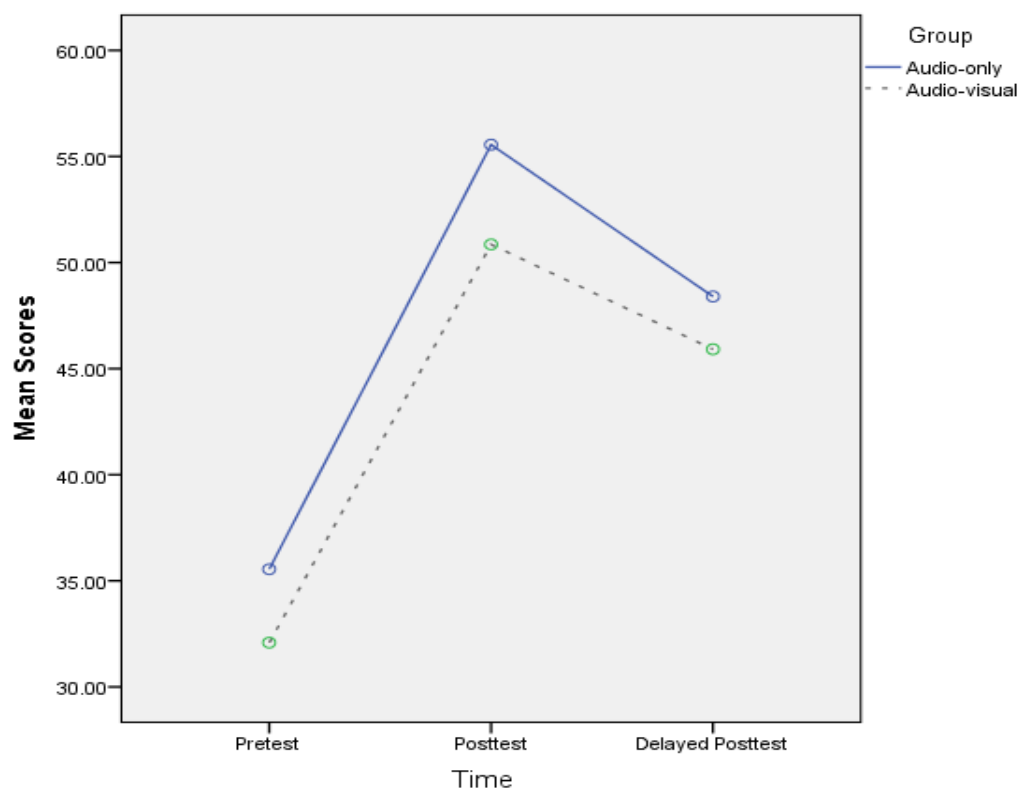


Figure 16. Production scores for pretest, posttest, and delayed posttest for AO and AV groups.

On the other hand, as table 15 reveals, there was a significant effect of time on the performance within groups,  $F(2, 56) = 35.03, p < .001$ . This result indicates that the production scores of one or both groups differed significantly between the pretest and delayed posttest. The test, nonetheless, does not reveal where exactly the difference lies.

To evaluate progress over time within each group, the students' mean scores on the pretest, posttest, and delayed posttest were compared for the AO group first, then the AV group. Table 15 provides the mean percentage scores and standard deviations for the three tests for both groups. In addition, it reports the results of the one-way ANOVA tests conducted on the two groups.

The results of the ANOVA test conducted on the AO group indicated that learners' performance differed significantly over time,  $F(2, 28) = 18.02, p < .001, \eta^2 = .56$ . Thus there is evidence that there was a change in participants' scores. This reveals that AO students' linking production changed significantly over time from pretest to delayed posttest.

Table 15. AO and AV one-way repeated measures ANOVA results

	Test	Mean Score	SD	Sum of squares	df	Mean square	F	Sig.
<b>AO</b>	Pretest	35.55	19.09	3083.22	2	1541.61	18.02	.000
	Posttest	55.55	22.76					
	Delayed	48.40	18.88					
<b>AV</b>	Pretest	32.09	20.75	2838.84	2	1419.42	17.10	.000
	Posttest	50.86	13.49					
	Delayed	47.91	15.59					

To determine which tests were significantly different from each other, a pairwise comparison showed that there was a significant increase in AO scores over time from the pretest to the posttest and from the pretest to the delayed posttest. The participants' mean percentage score was 35.55 at pretest, 55.55 at posttest and 48.40 at delayed posttest. The pairwise

comparisons revealed that students' change scores between pretest and posttest were significantly different,  $t(14) = -5.76, p < .001$ , and between pretest and delayed posttest were also significantly different,  $t(14) = -3.42, p = 0.012$ . Nevertheless, students' change scores between posttest and delayed posttest were not significantly different,  $t(14) = 2.52, p = 0.073$ . It can be concluded that students' proficiency in linking production significantly improved immediately after training with AO feedback, and decreased when evaluated one month after training. However, their delayed posttest scores were still significantly greater than their pretest scores.

Similarly, in examining the overall time effect on the AV students' production scores, the results of the one-way repeated measures ANOVA indicated that students' performance differed significantly over time,  $F(2, 28) = 17.20, p < .001, \eta^2 = .55$ . This reveals that AV students' linking production changed significantly over time from pretest to delayed posttest.

The pairwise comparisons showed that there was a significant increase in AV scores over time,  $p < .05$ , from the pretest to the posttest and from the pretest to the delayed posttest. The participants' mean percentage score was 32.09 at pretest, 50.86 at posttest, and 45.92 at delayed posttest. The pairwise comparisons showed that students' change scores between pretest and posttest were significantly different,  $t(14) = -5.60, p < .001$ , and between pretest and delayed posttest were also significantly different,  $t(14) = 4.13, p = .011$ . Yet, students' change scores between posttest and delayed posttest were not significantly different,  $t(14) = 3.39, p = .144$ .

To conclude, the results indicate that students who took AO and AV linking training dramatically improved their production of linking in the text reading task immediately after training. AO and AV students were also able to retain the majority of this improvement when evaluated one month after training.

### **Research question 3: Transfer to novel contexts**

To answer the third research question as to how improvement in linking production transferred to novel contexts, I compared students' mean scores on the 'old' part of the posttest to the 'new' part that was added in the posttest. A paired sample *t*-test was conducted to evaluate whether a statistically significant difference existed between the mean scores of the old and new parts of the posttest for the AO group then the AV group. Table 16 summarizes the differences

between the two tests, the difference values, the  $t$ -values and their degrees of freedom ( $df$ ), and the significance of the tests ( $p$ -value) of the two groups.

Table 16. Results of paired  $t$ -tests on production posttests for old and novel contexts

	<b>Context Novelty</b>	<b>Mean score</b>	<b>SD</b>	<b>Mean difference</b>	<b><math>df</math></b>	<b><math>t</math></b>	<b>Sig.</b>
<b>AO</b>	Old	55.55	22.76	10.63	14	2.21	.045
	New	44.92	17.43				
<b>AV</b>	Old	50.85	13.49	4.92	14	1.87	.082
	New	45.92	15.51				

The results of the paired sample  $t$ -test (Table 16) were significant for the AO group,  $t(14) = 2.21$ ,  $p = .045$ ,  $\eta^2 = .25$ , indicating that there was a significant difference between the posttest old score and the posttest new score. The production improvement that the AO students achieved while reading the old text decreased from  $M = 55.55$  to  $M = 44.92$  in the new text. The mean decrease was 10.635, with the 95% confidence interval for the difference between the means of .29 to 20.97.

On the other hand, the results of the paired sample  $t$ -test were not significant for the AV group,  $t(14) = 1.87$ ,  $p = .082$ ,  $\eta^2 = .19$ . The production improvement that the AO students achieved while reading the old text decreased from  $M = 50.85$  to  $M = 45.92$  in the new text, yet not significantly. The mean decrease was 4.92, with the 95% confidence interval for the difference between the means of -.71 to 10.57. Figure 17 shows the difference between the scores of the old and new parts of the posttest for the two experimental groups.

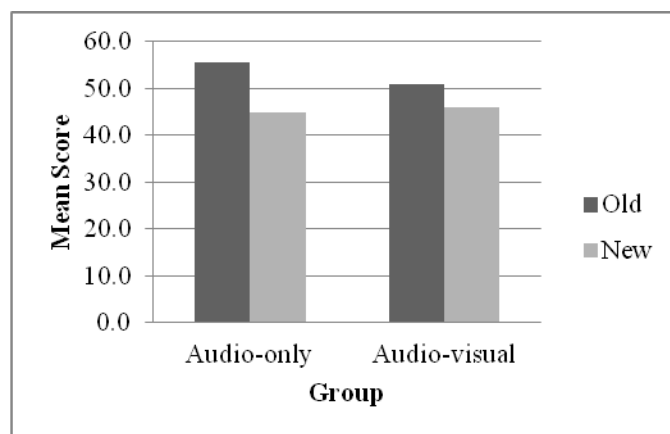


Figure 17. Production scores for posttest old and new parts for AO and AV groups.

In general, AO and AV participants improved their linking on the new part less than they did on the old part of the reading text. However, AV training helped to significantly transfer the gained improvement to novel contexts better than the AO training.

#### Research question 4: The effect of word frequency

To answer the first part of the fourth research question as to how word frequency impacted linking production, a paired *t*-test was conducted to compare participants' production of linking to/from high frequency words to the production of linking to/from low frequency words in the posttest. Results showed that LF words ( $M = 21.82$ ) were more difficult to link to/from than HF words ( $M = 53.20$ ),  $t(29) = 12.68$ ,  $p < .001$ ,  $\eta^2 = .85$ . This result indicates that students link better in HF contexts than in LF contexts.

To answer the second part of the fourth research question as to whether learners' production improved better in HF contexts than in LF contexts after training, a paired *t*-test was used to compare the change of mean score between the pretest and posttest of HF and LF words. The test revealed that students' improvement in linking HF words was significantly greater than that of LF words,  $t(29) = 4.02$ ,  $p < .001$ ,  $\eta^2 = .35$ . Figure 18 shows differences in improvement in HF and LF linking over time.

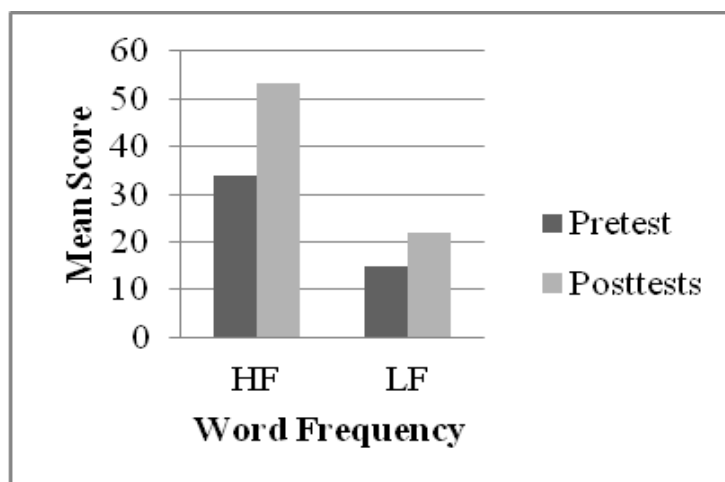


Figure 18. Mean pretest and posttest scores for HF and LF words

Students were able to significantly increase the mean scores of linking HF words by 19.39 percent,  $t(29) = -7.99$ ,  $p < .001$ , while increasing the mean scores of linking LF words by 6.99 percent,  $t(29) = -2.72$ ,  $p = .011$ . Table 17 provides the mean scores and standard deviations for students' linking production of HF and LF words, in addition to the results of the paired  $t$ -tests performed on the scores of each text separately.

For the fourth research question, the results indicate that high frequency words were easier to link for NNSs than low frequency words. They also demonstrate that students were able to attain a higher level of improvement in linking when they read a text comprised of HF words than when they read a text that had a high percentage of LF words.

Table 17. Results of paired  $t$  test for HF and LF word linking

Word Frequency	Test	Mean score	SD	Mean difference	df	t	Sig.
HF	Pretest	33.81	19.67	19.39	29	-7.99	.000
	Posttest	53.20	18.54				
LF	Pretest	14.83	16.54	6.99	29	-2.72	.011
	Posttest	21.82	18.32				

## Discussion

This section provides a summary and discussion of the major findings in the production part of the study. It analyzes the results in the light of relevant previous research. In addition, the section closely examines examples from the data to provide insights on the phenomenon of linking and the factors that could affect learners' production of linking. The section is organized by research questions.

### Research question 1: The effectiveness of training

The first research question investigated how effective AO and AV training were in improving linking production for non-native learners of English. To determine the effectiveness of each training approach, 42 students took a text reading task where they read a short story before and after training. The students participated in a two-week training that involved four 50-minute sessions during their regular classes. The training included in-class and online materials with a variety of activities to improve students' perception and production of English linking. During the online training, students received either audio-only or audio-visual feedback on their performance depending on the group they belonged to.

Results of a one-way ANOVA indicated that there was a significant difference in CG, AO, and AV participants' production scores in the pretest and posttest,  $p < .05$ . Pairwise comparisons revealed statistically significant differences between the mean change of scores of the CG ( $M = 3.70$ ) and AO group ( $M = 20$ ), and between the CG and AV groups ( $M = 18.76$ ). The linking production of both experimental groups improved significantly more than the control group after training. The AO group performance improved slightly more than that of the AV group.

The finding that AO and AV instruction positively improved L2 learners' proficiency in linking agrees with previous studies on linking production (Kuo, 2009; Melenca, 2001; Sardegna, 2011). To explore the influence of explicitly teaching Japanese speakers of English how to connect speech, Melenca recruited four students in the control group and five in the experimental group. Their ability to link word pairs was rated using reading aloud and elicited free-speech monologues that were compared to a NS baseline. After three one-hour sessions of



training, the researcher found that the performance of the experimental group participants either improved, although not significantly, or remained relatively stable in linking ability while the CG performance stayed the same.

In a 14-week training program, Kuo (2009) examined the efficacy of teaching features of linking on the perception and production of elementary school Taiwanese students. The experimental group significantly improved their speech production and developed phonological awareness. Among the taught categories, V-V linking posed more problems for the experimental group due to its high degree of variance.

Sardegna (2011) trained 38 international graduate students on how to improve their ability to link sounds within and across words (among other pronunciation features that the study examined). A read-aloud test was administered and recorded twice during a one-semester course, and again five months to two years after the course ended. Findings of the study revealed that students' performance improved significantly after the training and was maintained over time regardless of the participants' native language, gender, and length of stay in the US prior to instruction.

Similar to previous studies, students in the experimental groups of the present study significantly improved their ability to link words. This result suggests that AO and AV training were effective in increasing students' accuracy in producing linking. Despite the relatively short period of training (four 50-minute sessions over two weeks), the study was able to achieve improvement levels similar to those in Kuo (2009) and Sardegna (2011) that trained students over a much longer periods of time (14-15 weeks).

The AO group slightly outperformed the AV group in linking production. This result contradicts previous research that found that learners who received both audio and visual feedback performed significantly better than those who received audio feedback only (de Bot, 1983; Hardison, 2004; Motohashi-Saigo & Hardison, 2009). One possible interpretation of the contradictory results is differences in students' learning styles. Although information processed through more than one cognitive channel (AV) is likely to be retrieved more successfully than when processed through only one channel (AO) (Paivio, 1971, 1991), the presence of visual feedback may have interfered with the ability of auditory-oriented learners to process linking.

This is especially true for students who were more used to auditory training in their previous L2 education. Such lack of readiness to deal with more than one channel of information requires more time for processing the two channels and more training in how to understand and utilize the visual feedback. The small sample size of participants in each group also limits the ability to interpret and generalize the results.

### Characteristics of NNSs' Linking

In order to get additional insights into how NNSs link their words, I took a closer look at the participants' production data as well as the production of the NSs in the baseline group whose speech was used in developing the grading criteria. This analysis will help in identifying factors that contributed to failure in linking. Instead of linking words, NNSs tended to insert a glottal stop before vowel-initial words, where the link should have taken place. On the other hand, the NSs occasionally produced a less audible glottal stop when they did not link, which was, however, much less frequent than NNSs.

One of the factors that contributed to 'unlinking' words was careful speech. When reading the fiction text, several students took care to enunciate almost every word fully. In this disfluent style of reading, the majority of words were stressed and most initial vowels started with a glottal stop which resulted in fewer links. Avoidance of reading errors could be the motivation for careful speech, but it seems that more Chinese learners than others failed to link because they "pronounce one word at a time in English just as they do when speaking Chinese" (Lin, Fan, & Chen, 1995, p. 9). One student in the AV group commented that in China he was taught that this (careful word-by-word speech) was the proper way to speak and read both in Chinese and English.

Another reason why participants inserted a glottal stop and broke the link was stressing the second word in a linking pair for prosodic or pragmatic purposes. For example, 16 out of the 18 students who did not link *now I* stressed the word *I* in *now I will surely die*, initiating it with a glottal stop and breaking the link with the previous word *now*. In the context of the story, they wanted to emphasize the pronoun *I* which would help point at the boy as the victim of the snake. On the other hand, only one of the five NSs in the baseline group stressed *I* and did not link it in

the utterance. Expressing emphatic meaning, hence, was realized not only by means of extra pitch prominence and duration, but also by inserting a glottal stop before initial vowels.

The majority of the incidents of unlinking seemed to be unintentional where students did not and could be attributed to a number of reasons among which are: lack of or insufficient knowledge of the rules of linking, stumbling over speech, L1 interference, phonological environment, and word frequency and difficulty level. Longer utterances or those containing multiple links were more challenging than others. They seemed to initiate a domino effect so that when a student stumbled over the first part of the utterance, s/he continued to read the rest of it in a choppy manner without linking any of the words. This was true even in the case of students who demonstrated good linking skills before they stumbled. Stumbling occurred due to the existence of unexpected words, difficult LF ones, longer utterances, and a variety of physical and psychological reasons such as fatigue and lack of confidence. For example, 13 students said or started to say *carried it to his home* instead of *carried it to its home* because they were more used to the former one (the use of the animate pronoun *his* to refer to a snake instead of the inanimate pronoun *its*). When they noticed their mistake, they fixed it and stumbled over the next part of the sentence(s) breaking all subsequent links.

The presence or absence of linking in participants' first languages could have also influenced their ability to perceive of and produce linking in English. L1 transfer plays a major role in accounting for the way in which L2 phonology deviates from the phonology of the target language (Altenberg & Vago, 1983; Broselow, 1984; Flege, 1980; Wu, 1993). Topping (1964) provided evidence that rules operating across words (such as linking) can be transferred into the L2, just like other types of phonological phenomena. Since linking can be transferred into the L2, some participants might have produced and understood English linking in correspondence with their native languages.

For speakers of Romance languages and other languages that contain forms of linking, the presence of the phenomenon in their languages is expected to facilitate the learning process. However, such an issue is not easily determined for Chinese students, who constituted the majority of the participants of this study. With lack of research on linking or resyllabification in their language, determining the influence of L1 transfer is not an easy task. However, L1 influence can be observed in patterns of pronunciation in the study. For example, twelve students

pronounced *before I* as /bɪfɔ̣aɪ/ where although they connected the two words, they deleted the final /r/. Consequently, those instances were not counted as linking. One of the reasons for these modifications is students' restricted phonemic inventory, which includes only certain vowels without off-glides (Melenca, 2001). Therefore, students may insert a glottal stop and/or use different sounds to compensate for what is missing in the phonological repertoire of their L1.

Another example of L1 influence on the linking of L2 learners is the case of the two Malaysian students. Glottalization of final stops is a widespread feature in the production of speakers of Malay English (San, 2010) which prevents the realization of linking between words. For example, in the pretest, both Malaysian speakers glottalized the final stops of the first words in the pairs *top of* /tʌʔəv/ and *it up* /ɪʔ ʔʌp/. Their linking rates in general were among the worst in the two experimental groups (M = 7.41) and (M = 3.70) respectively. However, both of them improved significantly after training (M = 40.74) and (M = 37.04) and were able to unglottalize the final stops and link properly three out of four instances of the aforementioned pairs.

A closer examination of the scores of certain linking pairs provides further insights into the effect of phonological environment and word frequency on the production of linking, as well as the learnability of certain features of linking. Table 18 includes percentage mean scores of selected linked words from the pretest and posttest and the difference between the two tests for combined means of AO and AV groups. It also includes the mean scores of the NSs baseline group (Appendix X includes more details about all linked words).

Results of the study are in line with previous research regarding the types of linking least amenable to training. Research showed that the CSPs least responsive to training were /h/ deletion, as in *e* for *he* and flapping (Crawford, 2006; Ting & Kuo, 2012). In pairs like *when he* and *asked him*, the majority of students pronounced the /h/ which prevented linking the final consonant with the vowel after /h/. Although the training included explicit training on linking with /h/ deletion, students showed very little improvement on this aspect of linking, which suggests that /h/ deletion complicates linking and may need extra investigation and practice.

Table 18. Item analysis of selected production mean scores

Items	Combined AO & AV			NSs
	Pretest	Posttest	Difference	
when he	3	7	3	80
snake asked	13	23	10	60
top of	43	83	40	100
boy answered	10	13	3	60
you up	33	67	33	100
me and	3	13	10	80
asked him	0	13	13	60
he agreed	50	62	20	80
top of (2 <sup>nd</sup> occurrence)	57	80	23	100
can I	97	97	0	60

In addition, V-V linking, with a larger degree of variance than other types of linking, poses more problems to L2 learners (Kuo, 2009). When reading *he agreed*, many Chinese students altered the pronunciation of the initial vowel in *agreed* in a way that resulted in a continuum of low central vowel variants similar to /hi ʌgrid/, /hi ʌgrid/ or /hi ɪgrid/ many times breaking the link between the two words. These variations happened due to the restricted phonemic inventory for Chinese students, so they attempted to compensate using vowel variants from their L1. V-V linking poses an additional challenge to L2 learners because of glide insertion. It was particularly difficult for Chinese students to insert the glide /j/ before similar sounds like /ɪ/. V-V linking appears to be challenging to learners, and may require either additional training or a different type of training.

Another difficult aspect for students was linking to reporting verbs in such pairs as *snake asked* and *boy answered*. Students' scores were low in the pretest for these items and improved slightly in the posttest by 10 percent and 3 percent respectively. It is not clear why very little improvement was attained on linking reporting verbs. However, it is possible that students, in

anticipation of the quote or nominal clause after the reporting verb, broke the link and stressed the verb to separate the two parts of the sentence. Looking at NSs' scores on the same items, it seems that such verbs were also associated with relatively lower linking scores (60 percent). Consequently, the linking of reporting verbs should probably not be emphasized when teaching linking due to the low rate of response to training.

On the other hand, certain high-frequency pairs that demonstrated high improvement rates should be prioritized when teaching linking. Those pairs were easier for students to link, such as *top of* (two occurrences) and *can I* (Table 18). Such items are usually stored as a single semantic and phonetic entity (Field, 2003b), so students are able to retrieve them quickly. Training was not effective on some of the very frequent chunks due to reaching a ceiling effect. For instance, 97 percent of the students linked *can I* correctly in the pretest which left no room for improvement in the posttest. Yet other high-frequency chunks were very responsive to the training, for example, *top of* improved by 40 percent after training.

The findings, coupled with the results of previous research, indicate that both AO and AV instruction are significantly effective in improving learners' production of linking. The discussion provided more insights into how NNSs produce linking. When they do not link, L2 learners insert a glottal stop intentionally or unintentionally at the beginning of the second word. They might break the link to pause and to express emphatic meanings. On the other hand, unintentional insertion of glottal stops could be attributed to a number of reasons including lack of or insufficient knowledge of the rules of linking, stumbling over speech, L1 interference, phonological environment, and word frequency.

### **Research question 2: The long-term effects of training**

The second research question investigated how linking production improvement caused by training with AO feedback or AV feedback persisted. To evaluate the long-term effects of the training, students took a delayed posttest one month after the end of the training. The delayed posttest was identical to the immediate posttest. The mean scores of the two experimental groups were compared using a two-way repeated measures ANOVA, followed by post hoc pairwise comparisons to identify significant differences. Results revealed no significant differences between the performance of the AO and AV groups over time. However, both groups retained a

significant part of the improvement they attained during training by the time of the delayed posttest. Although the performance of the AO group decreased by 7.16 percent, and the AV group performance declined by 4.94 percent one month after the end of training, both groups remained significantly above their pretest mean score.

The results agree with previous research that showed that learners' performance generally decreased upon discontinuing training, yet remained above their pretest average score (Ruellot, 2006; Sardegna, 2011). Sardegna's study, in particular, focused on evaluating the long-term effectiveness of covert rehearsal by analyzing students' progress in reading English text at different times from five months to two years. She found that students' performance improved significantly after the training and was maintained over time regardless of the participants' native language, gender, and length of stay in the US prior to instruction.

### **Research question 3: Transfer to novel contexts**

The third research question examined how improvement in linking production transferred to novel contexts that were not on the pretest. In order to evaluate the generalization of training improvement to linking new words, a second part was introduced in the posttest that included the rest of the fiction story. Since students had not read this part in the pretest, it was expected that if their scores improved significantly on this part, then training gains were successfully transferred to new contexts and the repetition of the first part of the test in the posttest did not interfere in their scores (i.e., no test effect).

The mean scores of the old and new parts of the posttest were compared using paired *t*-tests for the AO and AV groups. Results were significant for the AO group with a mean score decrease of 10.63 percent. This means there was a significant difference between the scores on the old and new parts of the posttest and, consequently, no significant improvement on new items for the AO group.

On the other hand, the results of the *t*-test were not significant for the AV group whose performance in the new posttest decreased by 4.92 percent. This minor decrease in scores translated into significant transfer of improvement to new contexts. Because AV students' performance on the two parts of the posttest was relatively similar, their prior knowledge of the first part of the posttest did not affect their performance and any improvement gains may be

attributed to the training. This claim cannot be validated in the case of AO group since there was a significant difference between the scores of the two parts of the test.

The results of the AV group are in line with previous research that supported the ability of learners to generalize improved perceptual capabilities to novel stimuli (e.g., Hardison, 2003; Hirata, 2004; Lively, Logan & Pisoni, 1993; Motohashi-Saigo & Hardison, 2009; Ruellot, 2006). AV training helped to transfer the gained improvement to novel contexts better than the AO training.

#### **Research question 4: The effect of word frequency**

The fourth research question examined the impact of word frequency on linking and the effectiveness of training on low frequency words. In order to evaluate learners' ability to link words of varying frequency levels, a second text was added to the pretest and posttest that included a high percentage of LF words. Comparing students' scores in the HF and LF posttest highlighted differences in their linking performance depending on words frequency. On the other hand, comparing students' scores in the HF and LF pretest and posttest provided data on the effectiveness of instruction (both AO and AV) on the production of linking HF and LF words.

In answering the first part of the fourth question, the results of a paired *t*-test demonstrated a significant difference between students' scores on the HF posttest words ( $M = 53.20$ ) and LF posttest ( $M = 21.82$ ). This indicates that LF words were significantly more difficult to link to and from than HF ones. Another *t*-test was conducted to compare the pretest and posttest scores of HF and LF words and revealed that students' improvement of linking HF words was significantly greater than that of LF words. In addition, the increase of mean scores due to training was significant for HF with mean score gains of 19.39 percent, and for LF with mean score gains of 6.99 percent. This suggests that HF words are easier to link for L2 learners and more responsive to training than LF words.

No previous research has been conducted on the difference between the behavior and instruction of HF and LF words in connected speech. However, the findings of the study are in line with general previous research on word frequency that proposed that high frequency words and phrases are more likely to undergo processes of phonetic reduction (Bybee, 2001; Ellis, 2002). Bybee (2001) noted that such reductions must be represented as part of the stored image



of HF words and phrases, which are, then, processed as units. This process was manifested in the data of the present study. Pairs that received higher percentage of linking were those containing more higher frequency words and used more frequently together, e.g., *can I, take it, threw it, pick up, many of*; in comparison to *me and, sat under, to its, them are*. Unlike the latter set of phrases, the words in the former set form lexical bundles and are often used together. The repetitive usage results in even more reduction and linking. This implication reinforces the significance of teaching high frequency phrases from the early stages of L2 education, especially that HF chunks are more amenable to instructions than LF ones. Not only are the learners' skills on producing the language heavily dependent on the number of words they know (Nation, 1990), but their speaking fluency develops better with HF phrases.

The influence of LF words on participants' linking production can be specifically observed in the data of the study. Because reading the LF texts was challenging for many participants, they employed different strategies to perform this task. Some read the text, stumbling over many of the LF words and eventually becoming frustrated with the rest of the text, while others unconsciously replaced several of the difficult LF words with words they were more familiar with (e.g., *accompany* for *accommodate*). Some participants could not link HF pairs that were followed by a LF word. For instance, one learner could not link *to a* because he was busy attempting to spell the next word *variety*. The majority of students broke the links because they were challenged by the LF words.

An analysis of NSs' production of the two reading texts showed that their linking mean score for the HF text was 79%, while that for the LF text was 87%. The results indicate that NSs performed the majority of links in both conditions, with an unexpected higher rate for the LF text. Because of the high concentration of academic LF words in the LF text, it was expected that NSs and NNSs alike would attain lower linking scores. On the contrary, NS linked more in the academic LF text, which suggests that the phenomenon of linking for NSs was not affected by word frequency. For NNSs, however, linking is influenced by word frequency. LF words, in particular, are a major cause of linking difficulty in connected speech production.

## Chapter Summary

This chapter examined the effectiveness of instruction on linking production using AO and AV feedback. First, it described the instrument used to collect production data, namely the pretest, posttest, and delayed posttest. Second, it explained how participants' production data were randomized and prepared for analysis, and defined the criterion for rating linking. It also described the statistical analysis scheme. Finally, the findings of this part of the study were discussed in light of relevant literature.

The results revealed a significant effect of instruction on L2 learners' production of linking for both groups immediately after and one month after the end of training. The AV training was successful in helping students transfer gained improvement to novel contexts. The same could not be stated about the AO group. High frequency words were found significantly easier to link and more responsive to training than LF words.

The chapter also analyzed specific examples from the study to shed some light on the characteristics of NNSs' linking. The researcher observed that when learners did not link, they inserted a glottal stop at the beginning of the second word. They broke the link to pause and to express emphatic meanings. Additionally, they sometimes inserted a glottal stop due to lack of or insufficient knowledge of the rules of linking, stumbling over speech, L1 interference, phonological environment, and word frequency.

## **CHAPTER VI: LEARNERS' PERSPECTIVES ON THE EFFECTIVENESS OF LINKING INSTRUCTION**

In addition to investigating the effectiveness of instruction on learners' linking perception and production, the study surveyed participants' overall impressions on AO and AV training. The purpose of this analysis was to help in interpreting the quantitative results and provide insights into the training benefits and limitations as a tool in the L2 classroom. This chapter describes the post-training questionnaire that was utilized to collect participants' perspectives and the methods used to analyze participants' responses. It furthermore reports the qualitative findings of the study and discusses them in light of the quantitative results and relevant literature.

### **Post-training Questionnaire**

In addition to providing biographical data in the background survey at the beginning of the study, participants completed a post-training questionnaire to gather their opinions about the training, its effectiveness, advantages, and limitations. The questionnaire included both open-ended and close-ended questions in order to employ a combined methodological approach for analyzing data using both quantitative and qualitative data (Dörnyei, 2009). Results of the close-ended questions suggested general trends in the participants' opinions, and answers to the open-ended questions may contribute to explaining the close-ended quantitative descriptions, as well as add depth to the major findings of the study.

The first section of the questionnaire included the close-ended questions: The AO version included 11 Likert scale questions while the AV questionnaire included 16 because it contained additional questions about the use of waveform in the training (see Appendix XI for post-training questionnaires). Some of the questions were modified depending on the type of training each group received so that they reflected the type of feedback used, as can be seen in Table 19.

For every question, students were asked to read a statement and indicate the extent to which they agreed or disagreed with it. Each question offered four response choices: agree, somewhat agree, somewhat disagree, and disagree. The questions revolved around five categories in order to present results in a clearer and more accessible manner. The categories

were: (1) ‘training satisfaction’ to evaluate the overall effectiveness of the training and the time allotted to such training, (2) ‘task features’ to rate the usefulness of several aspects of the training such as recording, comparing and authenticity of materials, (3) ‘teacher’s role’ to examine the extent of teacher’s support in conjunction with online materials that encourage learner’s autonomy, (4) ‘training outcome’ to examine perceived gains in linking perception and production, and (5) ‘EVF features’ to explore the perceived role of waveforms in improving linking. The last category was limited to the AV group. Table 20 shows the questionnaire categories and their corresponding Likert scale questions for AO and AV groups.

Table 19. Post-training questionnaire questions for AO and AV groups

Categories	AO Group	AV Group
Training satisfaction	Overall, I liked the past pronunciation training sessions.	Overall, I liked the past pronunciation training sessions.
	I would like more training similar to this training.	I would like more training similar to this training.
	The number of practice sessions was appropriate.	The number of practice sessions was appropriate.
Task features	Being able to record (and re-record) myself while practicing helped me improve my pronunciation.	Being able to record (and re-record) myself while practicing helped me improve my pronunciation.
	It was easy to compare my recording to the one on the website.	It was easy to compare my waveform to the one on the website.
	The examples taken from movies were easy to understand (Authenticity).	The examples taken from movies were easy to understand (Authenticity).
	I’d like to have more examples taken from movies (Authenticity).	I’d like to have more examples taken from movies (Authenticity).
Teacher’s role	The teacher’s explanation of connected speech helped me improve the way I use it.	The teacher’s explanation of connected speech and how to interpret waveforms helped me improve the way I use them.
	This online practice allows me to work independently on my listening and pronunciation.	This online waveform practice allows me to work independently on my listening and pronunciation.

Table 19 (Continued)

<b>Categories</b>	<b>AO Group</b>	<b>AV Group</b>
Training outcome	This training helped me notice how I connect words.	This training helped me notice how I connect words.
	I can understand connected words better in speech after this training.	I can understand connected words better in speech after this training.
EVF features		The waveform activity was helpful for noticing how words are connected.
		I found it helpful to compare the waveform of the model speaker with my own.
		It was easy to see the link between words in the waveform.
		Writing the words under each waveform helped me understand it.

The second section of the questionnaire included five open-ended questions that were designed to collect more detailed accounts of students' perspectives of the training sessions. The first question asked students to reflect on the progress of their linking performance after the training. The next four questions asked the participants to provide feedback on the training indicating the features they liked the best/least, and to pinpoint any problematic or confusing aspects in the training. Finally, the questionnaire solicited suggestions from the students on how to improve the training further.

### **Analysis**

The fifth research question in the study examined learners' perceptions of the use of audio-only and electronic visual feedback in teaching linking. To answer this question, 29 of the 30 participants in the experimental groups filled out the questionnaire: 14 in the AO group and all 15 students in the AV group. Questionnaire responses were analyzed separately for each group using the statistical software SPSS. For Likert-scale questions with four response choices,

differential weights of 1 to 4 were assigned, with 1 being (disagree) and 4 being (agree). Then all student answers were entered into an Excel spread sheet and later imported into SPSS to calculate descriptive statistics for each survey item (mean and standard deviation). In addition, a Mann Whitney U test was run on every item to compare the answers of the two groups.

For students' narrative responses to the open-ended questions, responses were listed and grouped into categories (Dörnyei, 2009). In this process, key statements in every question were highlighted, and then corresponding responses were underlined and categorized in a similar fashion. A descriptive account of the most common and interesting responses is provided and discussed in the following sections.

## Results

The questions in the post-training questionnaire were grouped into five categories to present results in a clear thematic manner. The categories are: training satisfaction, task features, teacher's role, training outcome, and EVF features. The last category is limited to the AV group. Table 20 shows the questionnaire response means for AO and AV groups as well as their standard deviation. It also includes the significance values calculated by Mann Whitney U tests to compare the ratings of the two groups.

Table 20. Means and standard deviations of AO and AV group perceptions of linking

Categories	Questions	AO group Mean	SD	AV group Mean	SD	<i>p</i>
Training satisfaction	Overall, I liked the past pronunciation training sessions.	3.14	.91	3.47	.44	.16
	I would like more training similar to this training.	2.75	1.09	3.07	.86	.36
	The number of practice sessions was appropriate.	3.46	.95	3.17	.82	.13

Table 20 (Continued)

Categories	Questions	AO group Mean	SD	AV group Mean	SD	<i>p</i>
Task features	Being able to record (and re-record) myself while practicing helped me improve my pronunciation.	3.14	.91	3.27	.88	.63
	It was easy to compare my recording (waveform) to the one on the website.	2.86	1.03	3.43	.62	.07
	The examples taken from movies were easy to understand (Authenticity).	3.25	.98	3.63	.61	.32
	I'd like to have more examples taken from movies (Authenticity).	2.79	1.17	3.57	.50	.01
Teacher's role	The teacher's explanation of connected speech (of how to interpret waveforms) helped me improve the way I use them.	2.75	1.09	3.47	.64	.02
	This online practice (waveform practice) allows me to work independently on my listening and pronunciation.	3.43	.81	3.73	.26	.02
Training outcome	This training helped me notice how I connect words.	3.25	.70	3.47	.64	.26
	I can understand connected words better in speech after this training.	3.04	.89	3.40	.60	.19
EVF features	It was easy to see the link between words in the waveform.	-	-	3.47	.44	-
	The waveform activity was helpful for noticing how words are connected.	-	-	3.53	.48	-
	I found it helpful to compare the waveform of the model speaker with my own.	-	-	3.47	.44	-
	Writing the words under each waveform helped me understand it.	-	-	3.60	.51	-

Scores out of 4

Concerning 'training satisfaction', both groups were satisfied with the training sessions; however, the participants of the AV group had generally more favorable opinions about the training ( $M = 3.47$ ) than the AO participants ( $M = 3.14$ ). The AO participants stated that they

wanted more training similar to the audio-only training they took ( $M = 2.75$ ) and thought the number of training sessions was adequate ( $M = 3.46$ ), while the AV participants wanted more training similar to their audio-visual one ( $M = 3.07$ ) and thought they had enough sessions in this training ( $M = 3.17$ ) but to a lesser extent than AO group. Despite discrepancies in AO and AV participants' opinions, no statistically significant differences were found in this category.

The second category examined students' perspective of the helpfulness and ease of using some 'training features'. In all of the production activities, students were asked to read a phrase or a sentence while paying attention to certain linking aspects, then record it and listen to themselves and compare the recording to the available NS's recording. They were urged to repeat the process as many times as needed until they were satisfied with their performance. The students expressed that they found the process of recording their production to be helpful in improving their pronunciation with comparable averages of ( $M = 3.14$ ) for the AO group and ( $M = 3.27$ ) for the AV group. They further reported that comparing those recordings to the NSs' ones and/or the waveforms online was somewhat an easy process for the AO group ( $M = 2.68$ ) and an easy process for the AV group ( $M = 3.43$ ). In this regard, the two groups differed, but not significantly ( $p = .07$ ). As for the feature of the authenticity of training materials, AO students ( $M = 3.25$ ) and AV students ( $M = 3.63$ ) agreed that the training stimuli taken from movies were easy to understand. However, AV students preferred to have similar stimuli ( $M = 3.57$ ) significantly more than the AO group ( $M = 2.79$ ),  $p = .01$ .

For the third category, 'teacher's role', students reported their opinion on the interacting role of the teacher and the online training materials in the classroom. The AV students deemed the teacher's explanation of connected speech significantly more helpful to understand linking and waveforms ( $M = 3.47$ ) than the AO group ( $M = 2.47$ ),  $p = .02$ . In addition, the online training (with waveforms) allowed the AV students to work significantly more independently on their listening and pronunciation ( $M = 3.43$ ) than the AO students ( $M = 3.43$ ),  $p = .02$ .

Regarding the fourth category, 'training outcomes', the two groups did not differ significantly on the perspective of the role of instruction in improving their perception and production skills. The training helped the AO students ( $M = 3.25$ ) and the AV students ( $M = 3.47$ ) to notice how they connected words. AO students ( $M = 3.04$ ) and AV students ( $M = 3.47$ )



believed that they were able to understand linked words better in speech due to this training. Hence, according to the participants, the training was successful in achieving its set goals.

Finally, only the AV group were asked to report their opinions on some features of the electronic visual feedback they received during training. The students believed that integrating waveforms as feedback in the online activities was helpful for noticing how words are connected ( $M = 3.53$ ). For them, seeing the link between words in the waveform was easy ( $M = 3.47$ ). After observing the link on Audacity, the participants stated that comparing their waveform to that of the model NS was very helpful to them ( $M = 3.47$ ). The whole process was manageable to them ( $M = 3.60$ ) partly because of the availability of text annotations that accompanied the waveforms and helped students interpret them and use them more effectively.

For the open-ended questions, students interacted positively with the prompts supplying a variety of suggestions. As a result of the training, seven out of the eleven students in the AO group who answered this question indicated that they noticed that their pronunciation improved. All AV students, except one, agreed that their training helped them notice how words were linked together, and helped them improve their listening and pronunciation. Two students who recognized their recent awareness of the process of linking reported that it was hard for them to change their accent and the way they speak.

Commenting about potential problematic aspects of the training, five AO students found no confusing aspects in the training, while two wanted a longer training and three doubted the goals of the training because, for them, people do not speak like that in real life. One of the AO students stated “linked pronounce! is it really used usually in daily life”. On the other hand, eight of the AV students claimed there were challenging aspects in the training, such as insufficient training time and difficulty in comparing waveforms.

When asked to identify their favorite aspect of the training, out of the ten AO students who responded, three said they benefited from the pronunciation practice, two found using the lab exciting, two liked the fact that they did not have homework and one student enjoyed the audio taken from movies. For the AV group, eight participants favored the activities performed to practice linking production and perception. As for the materials used, four students commended the use of waveforms to help “compare your wave with the online wave” and two

students enjoyed having movies integrated into the training items. As for the aspects students enjoyed the least, the AO students complained about the great number of exercises and the lack of fun activities. The AV participants agreed with these claims and two students added that comparing waveforms was challenging for them.

Lastly, to further improve the training sessions, students' suggestions centered on time of training, materials, and activities. More than 30 percent of the students wanted longer training to allow for extra time for practice with a possibility of offering the whole training as an online class. They asked for more examples and longer reading practice. Several students suggested the use of more "fun activities" such as "chatting and laughing" or chances to "speak together" with other students. Two students wanted to explore other ways of using waveforms to practice their pronunciation. One student summarized these suggestions:

The number of practice sessions are not enough, in the class, most students are not brave enough to practice the pronunciation, so actually we can't improve so much as we want. In addition, 40 minutes class time is so short that we can't practice more. also, we will forget some knowledge after the class.

## **Discussion**

The questionnaire results indicated that the participants in the AO and AV groups had an overall favorable opinion of the training, with the AV group expressing more positive views of their training. They welcomed the use of new approaches in linking perception and production training and expressed a need for more similar training.

More than half of the participants believed that the training was helpful in improving their linking perception and production. One AV student noted that "After I learn this lesson, I notice that when I watch the American Films, I can understand more." However, several participants doubted that their pronunciation improved because they believed they needed more practice time. Another reason was related to difficulties in comparing students' audio and waveform to the NS model. Consequently, it was difficult for several participants to use AO or AV feedback to correct their own production. On the other hand, input modifications, such as annotated waveforms, helped students interpret EVF, as one student expressed, "I liked the most

the way that the sentence is divided as a waveform which it helped me to imagine it when people talk, so I wouldn't need to use the program always to see the waveform.”

The AV group mean of the category ‘teacher’s role’ was significantly higher than that of the AO group. This result can be attributed to the nature of the audio-visual training itself. Using waveforms to practice linking required extra time to train participants on how to use and interpret this type of feedback. Although AV students did not get more training time than the AO students, they recognized the importance of teacher explanation and sufficient preparation to take full advantage of the training materials. The AV students, furthermore, seemed to have developed their autonomous skills of working on the online materials independently, which can also be attributed to the presence of AV materials in their training. One of the students in this group stated, counting the benefits of that in the online training,

We can do things as fast or as slow as we like or want or need. We do not have to do every activity with the rest of the class. If I can not do some exercise or I am struggling with it I can do it and re-do it as many times as I want until I understand.

Other factors that could have been important in the differences between the two groups are the varied motivation levels of students. It is true that the experimental treatments of the study were assigned randomly to the two classes, but by informally observing the two classes, I concluded that the AV group had a greater number of more motivated students than the AO group. Several students in the AO group felt that they did not have to take the required listening class in which the study took place, and, consequently, were somehow indifferent to the training.

The use of authentic movie materials was a controversial topic in the training. While a few students found them difficult, other students, especially in the AV group, welcomed the challenge as a way to prepare them for real-life language. Students also suggested adding more interactive materials to the training to make it more engaging and simulate real-life conversations.

The questionnaire findings support and help us better understand the quantitative findings of the study. In the questionnaire, the AV group expressed more favorable opinions on the training than the AO group. The same result was found for most of the research questions of the study where the AV participants attained slightly higher scores than the AO participants. The

findings are also in line with previous research on learners' perspectives on using EVF in the language classroom. Similar to the present study, most participants thought that EVF was an effective approach to improving perceptual awareness and pronunciation, and had the potential to be successfully implemented in the second language classroom (Anderson-Hsieh, 1992; Coniam, 2002; Hardison, 2004; Lara, 2009; Motohashi-Saigo & Hardison, 2009; Ruellot, 2006).

## **Chapter Summary**

This chapter explored L2 learners' perspectives on the effectiveness of instruction on linking perception and production. It started by describing the post-questionnaire, the data collection instrument to answer the fifth research question of the study. The questionnaire included both open-ended that collected detailed accounts of students' opinions and close-ended questions that revolved around five categories: training satisfaction, task features, teacher's role, training outcome, and EVF features. This was followed by a quantitative and qualitative analysis of the results of the questionnaire. The discussion section provided insights about students' opinions in relation to their performance in the study and relevant literature.

## **CHAPTER VII: CONCLUSION**

The final chapter starts by summarizing the major findings of the study for both the perception and production parts. It then describes the pedagogical implications of linking instruction in pronunciation and perception. Finally, the chapter concludes by discussing limitations of the study and exploring paths for future research in the field of connected speech perception and production.

### **Overview of the Study**

The study investigated the effectiveness of using two approaches to teach linking to NNSs of English. Audio-visual feedback, integrating both audio and electronic visual feedback, and audio-only feedback were employed in the development of online materials to help improve learners' perception and production of consonant-to-vowel and vowel-to-vowel linking. The long-term effectiveness of the instructional materials was also examined to see whether learners were able to retain the improvement beyond the training period. In addition, the study explored whether improvement transferred to novel contexts. The influence of high frequency and low frequency words on linking production was also examined. Finally, students' perspectives of using AV and AO feedback were reported and discussed to guide future implementation of the materials.

The study took place in three intact classes where every class was randomly assigned a type of treatment: the control group (N = 15) received no treatment, the AO group (N = 15) received training with audio-only feedback, and the audio-video group (N = 15) received training with audio and waveforms feedback. Students in the two experimental groups took a two-week linking training, while CG participants did not receive any training. At the beginning of the training, all participants completed a background questionnaire and took a perception and production pretest. After training, all participants took a posttest, then an identical delayed posttest one month after the end of training. In addition, AO and AV students completed a post-training questionnaire to collect their opinions of the training and its features and outcomes. Text and audio data were analyzed to answer the research questions posed by the study. The following section summarizes the major findings of the study.

## Summary of Major Findings

Since the study examined the effectiveness of instruction on linking perception and production, in addition to learners' perspectives on the training, separate chapters were dedicated to discussing every one of these themes. The first four research questions were answered for linking perception first, then linking production. Table 21 outlines the first four quantitative research questions of the study, together with their major results and significance levels.

Table 21. Summary of quantitative findings

	<b>Research question</b>	<b>Treatment</b>	<b>Effectiveness of training</b>	<b>Significant?</b>
<b>Perception</b>	Short-term improvement (pretest-posttest)	AO	Greater improvement than CG	No
		AV	Greater improvement than AO and CG	No
	Long-term improvement (pretest-delayed posttest)	AO	Increase from pretest to posttest Increase from posttest to delayed posttest Significant increase from pretest to delayed posttest	Yes
		AV	Significant increase from pretest to posttest Decrease from posttest to delayed posttest Increase from pretest to delayed posttest	Yes
	Transfer to novel contexts	AO	Significant transfer of improvement to novel contexts	Yes
		AV	Significant transfer of improvement to novel contexts	Yes

Table 21 (Continued)

<b>Production</b>	Short-term improvement (pretest-posttest)	AO	Significant improvement more than CG Slightly greater improvement than AV	Yes
		AV	Significant improvement more than CG	Yes
	Long-term improvement (pretest-delayed posttest)	AO	Significant increase from pretest to posttest Minor decrease from posttest to delayed posttest Significant increase from pretest to delayed posttest	Yes
		AV	Significant increase from pretest to posttest Minor decrease from posttest to delayed posttest Significant increase from pretest to delayed posttest	Yes
	Transfer to novel contexts	AO	No significant transfer of improvement to novel contexts	No
		AV	Significant transfer of improvement to novel contexts	Yes
	Word Frequency		HF words significantly easier to link than LF HF words significantly more responsive to training than LF words	Yes

### The effectiveness of training

To determine the effectiveness of training on linking perception, students took a sentence dictation test in the pretest and posttest. Results revealed that although the mean scores of the

AO and AV groups increased more than the CG after training, no significant differences were found between the three groups. The performance of the AV group was a little higher than that of the AO group. The insignificant results can be attributed to the short period of training, sample size of participants, and potential factors affecting the ease and learnability of linking perception. Such factors include frequency of linked words, their phonological environment and the presence of other CSPs at the same link.

To determine the effectiveness of training on linking production, students read a short story in the pretest and posttest. Results indicated a significant difference between the three groups where the performance of the AO and the AV groups increased significantly more than that of the CG. The learners broke the link to pause and to express emphatic meanings. They unintentionally inserted glottal stops and broke the link possibly due to a number of reasons including lack of or insufficient knowledge of the rules of linking, stumbling over speech, L1 interference, phonological environment, and word frequency.

#### Perception vs. production

In general, students in the AO and AV groups demonstrated much greater improvement in their production scores than their perception scores. This finding contradicts most previous research that found that perceptual learning transferred to improved production even in the absence of explicit production instruction. In the present study, improvement in production exceeded that in perception and did not seem to be connected to it. The discrepancy between perception and production scores calls for more research to understand why students' linking production improved much more than their linking perception despite taking the same training.

One possible reason is the interaction of different factors that could have affected students' perceptual skills. During the dictation task, students had to perform several tasks simultaneously: (1) perform a top-down analysis of the aural message and understand its message; (2) perform a bottom-up analysis and understand individual words; (3) isolate distractors, such as outside noise; (4) and transform the aural message into a written one, paying attention to spelling. Each task contained other smaller tasks. For example, in performing bottom-up analysis, the presence of multiple CSPs at word boundaries, such as C-V linking and /h/ deletion in *in his*, may have made it difficult for learners to understand the two easy words *in*



and *his*. On the other hand, the task of text reading seemed less challenging in the sense that students had fewer tasks to perform at the same time. The availability of a read text reduced their cognitive load so students did not have to create their own language as required in spontaneous speech. This all made learning linking perception more challenging than production. The same training could result in smaller improvement when measured using spontaneous speech.

The results of the study show that the perception of linking was more challenging than its production for NNSs. This suggests that perception improvement may need a longer period of training with high frequency chunks that are more responsive to training than others. They can gradually be followed by less frequent chunks once students start to internalize the rules of linking. Students should also be gradually exposed and trained to decipher chunks that contain a combination of CSPs that approximate the complexity of real running speech.

Another possible reason could be differences between students' starting perception and production proficiencies. While pretest perception mean scores ranged between 62-64 percent for all participants, those for the production pretest ranged between 32-35 percent. Students had more room for improvement in production than they did in perception. A more challenging perception test could result in a different outcome.

### **The long-term effects of training**

In order to examine the long-term effects of the training on linking perception, students took a delayed posttest one month after the end of the training. Results revealed that there was no significant difference between the scores of the two groups over time. However, both AO and AV training were effective in improving students' linking perception beyond the duration of the training. While the performance of the AO group continued to improve a little by the time of the delayed posttest, the AV group's performance declined slightly yet remained above the pretest mean score. This indicates that the students were able to sustain the improvement achieved over time. Production long-term scores were similar to perception. The scores of both AO and AV groups declined a little after the posttest but stayed above the pretest scores.

Students who took linking training were able to retain a significant part of their newly gained improvement in production and perception beyond the training. This is an indicator that the training achieved its goals, especially that improvement occurred after only a short period of

training, which suggests that a longer training may lead to even better performance and higher rates of retention of improvement. Many language teachers complain about lack of time to incorporate pronunciation training in general and connected speech training in particular (Rogerson, 2006). Consequently, teachers are encouraged to include linking training in their syllabi because short training was sufficient to render significant improvement in students' performance.

### **Transfer to novel contexts**

To determine the effectiveness of AO and AV training on novel perception contexts, students took a second part of the perception test during the posttest that contained new sentences not included in the pretest. Findings indicate a significant transfer of improvement to new contexts for AO and AV students. As for production, results showed no significant transfer of improvement on the new items for the AO group. On the other hand, the AV group was able to generalize the improvement they attained during training to novel contexts. Yet, differences in the performance of the AO group were approaching significance and did not lag far behind the AV group.

The results suggest that the training was effective by helping students recognize and produce linking in new contexts, which is the ultimate goal of any language teaching materials. They further indicate that the performance of the two groups on both tests reflected their actual knowledge of linking perception and production skills and was not the result of any test effect. The fact that students listened to the same sentences or read the same text three times did not influence their performance since they performed similarly on novel-item tests. Exposure to a variety of phonological contexts that include linking and other CSPs can increase learners' chances of recognizing such links in new contexts.

### **The effect of word frequency**

In order to evaluate learners' ability to link words of varying frequency levels, a second text was added to the production pretest and posttest that included a high percentage of LF words, in addition to the original HF text. Results revealed a significant difference between students' scores on the HF and LF posttest words. This indicated that LF words were

significantly more difficult to link to and from than HF ones. In addition, it was found that students' improvement of linking HF words was significantly greater than that of LF words. LF words seem to be a major cause of linking difficulty in the connected speech production of NNSs of English to the contrary of native speakers who performed more links in the LF text than in HF one.

Because many LF words are longer, more unfamiliar, or contain a challenging combination of sounds, they require more training and a higher language proficiency level in general. For this reason, it is encouraged that any training of linking perception and/or production starts by using stimuli containing HF words. Linking with lower frequency words can then be introduced gradually as students develop their linking and language skills. It is also recommended that linking training starts in the early stages of language learning when students start to acquire such HF words and chunks, and then continues as they accumulate more HF and LF words.

### **Learners' perception of the training**

To answer this question, students filled out a post-training questionnaire that included both close-ended and open-ended questions. Analysis of the data showed that the participants had an overall favorable opinion of the training, with the AV group expressing more positive views of their experience. They welcomed the use of new training approaches using waveforms, improved listening awareness and pronunciation, a greater sense of independence, and the use of authentic materials. Students also asked for longer training, more engaging materials, and the integration of similar training in their language classrooms.

Combined with the quantitative findings of the study, participants' opinions reinforced the important role that electronic visual feedback played in improving students' linking skills. Students in the AO and AV groups performed similarly with the AV group outperforming the AO group slightly in perception and the AO group attaining a minor improvement over the AV group in production. Nevertheless, the qualitative results showed that using waveforms in training increased students' interest in linking and encouraged them to spend more time and effort on the training. Creating a positive learning environment is an essential goal in any

language classroom. The integration of EVF with audio feedback was a step forward in that direction.

#### Audio-only and Audio-visual training

The results of AO and AV students were similar in many aspects in the study and different in a few, but the question yet to be answered is which method was more effective in improving students' linking perception and production. In terms of linking perception, the AV training was more effective in increasing students' scores immediately after treatment. In addition, AV students' improvement in novel contexts was greater than that of AO students. On the other hand, AO students' perception continued to improve by the time of the delayed posttest while that of the AV group declined, yet remained above the pretest score. The same was not true for linking production scores where the AO students' change scores increased a little greater than the AV students from pretest to posttest. Yet, the AV scores declined less than the AO's by the time of the delayed posttest. The AV students also had significantly transferred the gained improvement to novel production contexts while the AO students did not.

Although the improvement attained by the AV group was generally greater than that of the AO group, the difference between the two groups was not statistically significant. Previous research demonstrated that information processed through more than one cognitive channel is likely to be retrieved more successfully than when it is processed through only one channel (Paivio, 1971, 1991). However, in the present study and especially in linking production, the integration of two feedback types did not significantly differ from audio-only feedback. Such performance of the AV group could be attributed to the length of training and richness and novelty of the AV feedback.

AO and AV students followed the same training schedule that lasted for four 50-minute training sessions. However, the AV students had to dedicate more time of the training to learn about the use of waveforms, practice using them and work on extra activities in that regard. Meanwhile, the AO students did not have to work on these extra activities and had more time dedicated to practicing linking perception and production.

Another reason that could have affected the performance of the AV group was the richness of the information presented in the waveform that, despite all annotation and

simplification efforts, could have been overwhelming to students. Although the AV students were instructed on how to interpret waveforms and use them as a feedback source, working with two sources of feedback simultaneously in real time could have been overwhelming to students. This was especially apparent in regards to linking production where students, in addition to listening and interpreting the waveform, had to read the text. This could have resulted in slightly better scores for the AO group in linking production than the AV group. The novelty of using waveforms as visual feedback for speech could also have been a hindrance to more improvement for the AV group.

Practically speaking, it seems that the small increase in the AV group's performance over that of the AO group does not justify all the extra time and effort spent to train students on using waveforms and creating the materials themselves. Since the AV training was more effective in improving linking perception, its use can be more productive in raising students' awareness of linking, especially for students who have difficulty 'listening to themselves'. Such students explained that seeing a visual representation of their speech was eye-opening and helped them see what was actually happening in connected speech. AV training may result in more significant improvement if students are provided with longer training. However, this issue is left for future research. In sum, AV feedback provides an objective and automatic feedback that helps raise students' awareness of linking while listening to running speech, but using an audio-only feedback, especially when teaching linking production, can render comparable results with less efforts.

### **Pedagogical Implications**

The findings of this study have implications for language learners, language teachers, material developers, and researchers. They provide a glimpse of the positive effects that AO and AV instruction have on linking perception and production of L2 learners. This section will outline the status of teaching linking in textbooks. Then, it discusses the implications of the results of the current study in the light of learners' needs and priorities.

By reviewing the studies on connected speech processes, it is imperative to notice that despite the shortage of research on the instruction of CSPs, most pronunciation textbooks include a component that deals with such features in a variety of ways. Almost all ESL pronunciation

teaching materials address the issue of linking and, except for a few textbooks, they appear to agree on how to teach it using conventional repetition drills. This can be attributed to insufficient knowledge of the nature of linking, and impracticality and inadequacy of existent rules that fail to predict regular linking patterns in running speech. Some teaching materials (Grant, 2009; Lane, 2004; Sheeler & Markley, 1991) mention linking briefly as a secondary feature of other processes. The majority of textbooks, however, assign from one to several pages to teach linking to students (e.g., Dauer, 1992; Hewings & Goldstein, 1998). These pages are generally within chapters concerned with teaching connected speech modifications. In addition, they explain different types of linking (C-C, C-V, and V-V) ,providing examples and drills focused on repeating or identifying instances of linking.

A few other ESL pronunciation textbooks (Gilbert, 2001; Reed & Michaud, 2005) integrated linking repeatedly and consistently throughout the course of the textbooks in almost every chapter, but mostly in chapters dealing with the pronunciation of stops and continuants. Reed and Michaud considered linking to be their first and most important sound concept, an indispensable feature required to practice the other sound concepts, such as reduced, deleted, and contacted sounds. Linking is especially recycled through endings (-ed), and reduced forms of function words. For example, linking is recycled in an exercise on third person singular present tense endings (-s), where students were asked to listen and fill in the missing word(s).

Nevertheless, there is still an overwhelming need for language learning materials that integrate the teaching of connected speech in the syllabus of any language class. Indeed, a study that surveyed the perspectives of Taiwanese EFL teachers toward connected speech instruction (Rogerson, 2006) found that although 59% of the teachers were familiar with CSPs, many were uncomfortable teaching them. Teachers thought that the major reason for not teaching connected speech was insufficient and unsuitable materials. This implies that more authentic and innovative teaching materials on CSPs need to be developed and introduced to the classrooms to support teachers and students.

Findings of the present study will help fill this gap by providing AO and AV training that helped ESL students improve their linking perception and production of English. The improvement was sustained over time and was transferred to novel contexts, which suggests that it helped students continue to understand linked words and produce them. The audio-video

approach to teaching linking, in particular, emerged as more favorable to students, and one that helped them achieve slightly better results than the audio-only approach.

The integration of such an approach provides learners with tools that enable them to view language from a different angle as well as to work independently while receiving appropriate training. Because many teaching approaches encourage the teaching of citation forms of words, learners' expectations of what they will hear are sometimes influenced by exposure to the written language (Field, 2003b). Waveforms provide learners with a graphic representation of speech that allows them to view the difference between what they believe they hear and what is actually present in the speech signal. As some of the participants in the study expressed, this process can be "eye-opening" to them because they can finally understand what happens in speech and why they are not able to understand NSs.

The online training provided opportunities for learners to notice their errors together with immediate feedback upon completion. Students self-monitored their production and compared their recordings and/or waveforms to native speakers'. They were also able to redo any activity as many times as needed and work at their own pace. Such activities encouraged autonomous student work and bolstered learners' self esteem. Although L2 learners are in need of more pronunciation and bottom-up listening training, limited classroom time does not usually allow for such training. Online training with audio-visual feedback can hence fulfill these pedagogical needs in the classroom.

Another implication of the study is concerned with developing stimuli for teaching linking perception and production. The findings of the study suggest that teachers should emphasize the teaching of language as chunks, not as individual citation words. L2 learners in the study were able to understand and produce better linking when the potential links involved chunks such as *top of*, *carry it*, and *come over*. It seems that since learners acquire certain constructions as chunks, they tend to have fewer pauses (Goldman-Eisler, 1968) and keep them well-connected. For such an approach to be successful, teachers should start teaching language in chunks very early in the teaching process. If students are provided with linked examples of the target words on their first contact, they are likely to retain them in that manner. For example, instead of teaching students the meaning of the preposition *of* independently and pronouncing it in its citation form, the teacher can present it in a carrier chunk/phrase in its weak form (i.e.,

forms with weaker vowels like /ə/), such as *top of*. The citation form presents students with a strong form of the word starting with a glottal stop that students are not likely to hear in running speech. Presenting the word in a carrier phrase where linking is performed is an actual representation of how the word is spoken in real life. In fact, textbooks should introduce the weak forms of words as the primary variants, instead of the strong citation forms which can shape listeners' expectation of what they hear in the target language.

Unlike other CSPs, linking is a more pervasive phenomenon that can take place between any two words where the second word starts with a vowel. Therefore, the perceptual competence needed to recognize C-V and V-V linking can be different and more challenging to acquire than other CSPs that mostly involve a preconstructed set of high frequency reductions. It is easier for the listener to retrieve a preconstructed expression stored in the memory, such as *wouldja*, than to compute it (Ellis et al., 2008), which facilitates the recognition of such high-frequency expressions. However, because linking can occur at any C-V or V-V word boundaries, recognizing it in speech requires an additional set of skills. Being a rule-governed process can reduce the automaticity in the perceptual and production process.

A frequency-based approach to teaching linking can help reduce the cognitive load on the learner and increase automaticity. The higher the frequency of a construction, the more likely that its linking pattern will be preserved. In such an approach, linking pairs that have higher frequency rates are given priority in the classroom and language teaching materials. In other words, materials for teaching linking, and CSPs, should focus on chunks that are likely to appear together repeatedly in spoken language. The linking pattern of the two words is stored in memory and reinforced by frequent use. In this approach, for example, a teacher can incorporate teaching the linking pattern of *can I* simultaneously while teaching yes/no questions. In this way, students are introduced both to the right linking pattern between the two words and the reduced form of *can*.

Speech corpora are a good place to explore and locate such high frequency pairs. In this regard, function words play an important role in connected speech due to their high frequency and the high likelihood that they will undergo phonological change. Additionally, NSs link significantly more when words are linked to a function word than when they are linked to a content word (Alameen, 2007b). Another factor that helps learners in this respect is the faster



lexical access to function words, in contrast to content words, that allows them to remember these words with more ease (Segalowitz & Lane, 2000). This all suggests that a frequency-based approach to teaching linking will include many instances of function words. It also demonstrates that linking, as a process, cannot be taught in isolation from other CSPs, especially reduced forms of function words, deletion, and flapping. The presence of such processes at a C-V or V-V link can render it more complicated and even incomprehensible to learners. A frequency-based approach will incorporate all the frequently appearing function words in carrier phrases/chunks that reflect actual use in running speech. Integrating the teaching of linking with that of function words will give learners more opportunity to practice linking in context throughout the course, instead of having to apply rules on isolated sentences. In addition, learners will be able to relate various aspects of connected speech to each other, such as linking, reduced forms of function words, rhythm, and phrasing.

Finally, the study discussed the types of linking that were more amenable to improvement after training and, hence, more likely to be included in language learning materials. In a study that analyzed several coursebooks and pronunciation textbooks, Crawford and Ueyama (2011) found that textbooks were inconsistent in selecting what reduced forms to present, with only one form being found in over half of the books examined. This suggests that a consensus about which forms are important to teach has yet to develop. Findings of the present study indicated that V-V linking appeared to be challenging to learn due to its higher degree of variation. The process of /h/ deletion is another process that made linking more complicated for students to understand and produce. Both processes may require either additional or different type of training. Such findings should help teachers and textbook authors set priorities in teaching connected speech by starting to teach those forms more amenable to instruction, such as C-V, using a frequency-based approach.

### **Limitations of the Study**

It is difficult to make broad claims in language learning based on the findings of one study. Therefore, before drawing any recommendations for future research, it is important to consider some of the methodological limitations that may have affected the results of this study.

One limitation is the small sample size of the study, with fifteen students in every one of the three groups drawn from only one institution. Due to the number of participants, results of the study should be interpreted with caution. Further investigation with a larger sample from a greater number of educational institutions should allow for some preliminary generalizations to be made. This can be facilitated by the use of training online materials. Being available online with immediate integrated and student-generated feedback, this training can be disseminated to a larger number of participants, which in turn will help confirm the results of this study. The study sample also included a large number of Chinese native speaking students due to the high proportion of Chinese students in the international student population of Iowa State University. Such high concentration of one native language could have affected the results of the study due to L1 interference in students' performance. A more balanced group of students is desirable if future studies are to be conducted on this topic.

The duration of training is another limitation that needs to be taken into account. The training lasted for four 50-minute sessions over the period of two weeks. Results of previous research showed significant improvement after training that lasted for much longer periods of time (e.g., Kuo, 2009; Melenca, 2001; Sardegna, 2011). Participants in the present study were able to demonstrate significant gains in linking production and minor gains in perception after only two weeks of training. Offering longer training sessions or short sessions spread out over a longer period of time could yield very different and more positive results. Such training would provide more chances for extra practice and transfer of gained knowledge to novel contexts. It would also allow the AV participants more time to interpret the waveforms and utilize them more effectively.

### **Recommendations for Future Research**

Despite these limitations, findings from the present study provide a step forward in the field of connected speech instructional materials. With a small number of previous studies in the field, a need emerges to develop and evaluate existing and new linking and other CSPs instructional materials. Future research is needed to fully understand the effectiveness of using audio-only and audio-video approaches to teaching linking. Replication studies will add to the

small pool of literature dealing with the effectiveness of connected speech instruction on NNSs' perception and production.

Replication studies can also expand on this study by varying the investigation in different ways. First, they can investigate teaching other connected speech processes. The present study focused on teaching only C-V and V-V linking so as not to confuse the students, but other CSPs, such as deletion and flapping, are in need of more in-depth investigation as well. More research into other processes that establishes their significance to listening comprehension and intelligibility should help in setting priorities in the language classroom. This will help teachers and material developers focus on the processes most essential to students' needs at a specific age, proficiency level or for a specific purpose. For example, what CSPs should be included in a syllabus for a beginner-level listening class? And which stimuli and examples are most effective in teaching the target CSP(s)? As intelligibility is a more realistic goal for language learners than is native-like acquisition (Munro & Derwing, 1995), future research can explore CSPs that are more likely to affect intelligibility and, hence, deserving of more attention in the classroom.

Furthermore, the scope of future studies could be broadened to extend the time between the end of training and the delayed posttest. In this study, students took a delayed posttest one month after the end of training to examine how improvement in linking perception and production persisted. Most results indicated that students' scores decreased by the time of the delayed posttest yet remained above their pretest starting scores. However, the perception scores of the AO group continued to improve beyond the posttest and were significantly higher than their pretest scores. Future studies that investigate the effect of longer periods of time on the sustainability of improvement could help interpret this result.

Replicating the present study with participants from different language backgrounds might yield important results. This study and the majority of previous studies had mainly students who spoke Chinese, Japanese or Taiwanese as their L1. More research is needed to investigate the effectiveness of linking instruction on speakers from other language backgrounds. For example, the pretest scores of the two Malay speakers in the study were among the lowest of both groups, yet their scores improved significantly after training. Because glottalization of stops is an issue that can impede Malay speakers' intelligibility, linking training can be a very promising in improving their perception, production, and intelligibility.

Finally, future research can be conducted on students from different proficiency levels. Participants of the study were students in a university level ESL class that focused on teaching listening strategies. Their listening proficiency levels were similar at the onset of the semester. Conducting the same study with beginner-level students is important to evaluate the efficacy of the proposed frequency-based approach to teach linking. In this approach, the training includes many high frequency chunks that are likely to occur together repeatedly in spoken language. Such combinations of words are likely to preserve their linking pattern over frequent use. Because students learn or encounter these chunks in the early stages of their L2 education, it is essential that they acquire their linked patterns early and continue to practice them until they become part of their phonological repertoire.

## **Conclusion**

To conclude, this dissertation attempted to answer critical questions in the field of connected speech perception and production, a field that has been largely unexplored, particularly for linking. The study introduced the use of audio-visual training which was found to result in better improvement in linking perception and production as well as more positive learner feedback than an audio-only training. The study also provided valuable empirical data on the types of problems L2 learners encounter in connecting words and understanding connected speech.

Despite the limitations, the findings of the study have a number of important pedagogical implications for teachers, material developers and researchers interested in teaching connected speech and applications of electronic visual feedback. The study seeks to encourage language teachers to incorporate the teaching of linking and other connected speech processes in their language classroom. It, furthermore, encourages material developers to contribute to this research by further exploring and assessing the AV and AO training materials proposed by this study. Finally, it is hoped that the study carried out in this dissertation will encourage more much needed research in the field of NNS connected speech.

## REFERENCES

- Alameen, G. (2007). Audacity. *TESL-EJ*, 11(1), 1–11.
- Alameen, G. (2007). *The use of linking by native and non-native speakers of American English* (Unpublished MA Thesis). Iowa State University.
- Alameen, G., & Levis, J. (in press). Connected speech. In J. Levis & M. Reed (Eds.), *The Handbook of English Pronunciation*. Malden, MA: Wiley-Blackwell.
- Allerton, D. (2000). Articulatory inertia vs “systemzwang”: Changes in liaison phenomena in recent British English. *English Studies*, 6, 574–581.
- Altenberg, E., & Vago, R. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning*, 33, 427–447.
- Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmental. *System*, 20(1), 51–62.
- Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, 11(4), 5–21.
- Anderson-Hsieh, J. (1996). Teaching suprasegmentals to Japanese learners of English through electronic visual feedback. *JALT Journal*, 18(2), 315–325.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, 42, 529–555.
- Anderson-Hsieh, J., Riney, T., & Koehler, K. (1994). Connected speech modifications in the English of Japanese ESL learners. *IDEAL*, 7, 31–52.
- Audacity. (2013). Retrieved from <http://audacity.sourceforge.net/>. Last accessed on May 15, 2014.
- Barry, M. (1984). Connected speech: Processes, motivation, models. *Cambridge Papers in Phonetics and Experimental Linguistics*, 3, (no page numbers).
- Barry, M. (1985). A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics*, 4, (no page numbers).
- Barry, M. (1991). Assimilation and palatalisation in connected speech (pp. 1–9). Presented at the ESCA Workshop on Phonetics and Phonology of Speaking Styles, Barcelona, Spain.
- Beebe, L. M. (1988). The sociolinguistic perspective. In L. M. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives*. New York: Newbury House.

- Biber, D. (2009). *Register, genre, and style*. Cambridge, UK ; New York: Cambridge University Press.
- Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 5.3.51). Retrieved from <http://www.praat.org>. Last accessed on August 4, 2013.
- Bowen, J. D. (1975). *Patterns of English pronunciation*. Rowley, MA.: Newbury House.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /ɾ/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299–2310.
- Broselow, E. (1984). An investigation of transfer in second language phonology. *International Review of Applied Linguistics*, 22, 253–269.
- Brown, A. (2013). Phonetics and phonology: Historical overview. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Oxford: Blackwell Publishers.
- Brown, G. (1990). *Listening to spoken English* (2nd ed.). London; New York: Longman.
- Brown, J. D., & Hilferty, A. (1986a). Listening for reduced forms. *TESOL Quarterly*, 20(4), 759–763.
- Brown, J. D., & Hilferty, A. (1986b). The effectiveness of teaching reduced forms for listening comprehension. *RELC Journal*, 17, 59–70.
- Brown, J. D., & Kondo-Brown, K. (2006a). Introducing connected speech. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 1–15). Manoa; Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- Brown, J. D., & Kondo-Brown, K. (Eds.). (2006b). *Perspectives on teaching connected speech to second language speakers*. Manoa; Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Cahill, R. (2006). Teaching reduced interrogative forms to low-level EFL students in Japan. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 99–125). Manoa; Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- Carreira, J. M. (2008). Effect of teaching reduced forms in a university preparatory course. In K. Bradford-Watts, T. Muller, & M. S. Swanson (Eds.), *JALT2007 Conference Proceedings* (pp. 200–207). Tokyo: JALT.
- Cauldwell, R. (2013). *Phonology for listening: Teaching the stream of speech*. Birmingham: speechinaction.

- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. New York: Cambridge University Press.
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation paperback with audio CDs (2): A course book and reference guide* (2nd ed.). Cambridge University Press.
- Chapelle, C. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning and Technology*, 2(1), 22–34.
- Chapelle, C. (2001). *Computer applications in second language acquisition*. Cambridge, UK: Cambridge University Press.
- Chun, D. M. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal*, 7, 21–47.
- Clarey, M. E., & Dixon, R. J. (1963). *Pronunciation exercises in English*. New York: Regents.
- Clark, J., & Yallop, C. (1995). *An introduction to phonetics and phonology*. Oxford, England: Blackwell.
- Cobb, T. (n.d.). *Web Vocabprofile*. Retrieved September 28, 2013, from <http://www.lex tutor.ca/vp>. n adaptation of Heatley, Nation & Coxhead's (2002) Range. Last accessed on September 11, 2013.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 edition.). Hillsdale, N.J: Routledge.
- Cohn, A. C. (1993). Nasalisation in English: Phonology or phonetics. *Phonology*, 10(1), 43–81.
- Coniam, D. (2002). Technology as an awareness-raising tool for sensitising teachers to features of stress and rhythm in English. *Language Awareness*, 11(1), 30–42.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Crawford, M. J. (2006). A study on teaching reductions perceptually. In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), *JALT 2005 Conference Proceedings*.
- Crawford, M. J., & Ueyama, Y. (2011). Coverage and instruction of reduced forms in EFL course books. *The Language Teacher*, 35(4), 55–61.
- Cruttenden, A. (2008). *Gimson's pronunciation of English*. London: Hodder Education.
- Cutler, A. (1990). Exploiting prosodic possibilities in speech segmentation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing*. Cambridge: MIT.
- Dauer, R. M. (1992). *Accurate English: A complete course in pronunciation*. Prentice Hall.

- De Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, 26(4), 331–350.
- Dickerson, W. B. (1994). Empowering students with predictive skills. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New views, new directions* (pp. 17–33). Alexandria, VA: TESOL Publications.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press, USA.
- Dörnyei, Z. (2009). *Questionnaires in second language research: Construction, administration, processing*. Mahwah, NJ: Erlbaum.
- Ellis, N. (2002). Frequency effects in language processing: A review with implications of implicit and explicit theories of language acquisition. *Studies in Second Language Acquisition*, 24(2).
- Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396.
- Eskenazi, M. (1999). Using a computer in foreign language pronunciation training: What advantages. *CALICO Journal*, 16(3), 447–469.
- Fan, J. Y. (2003). The effect of teaching connected speech rules on English listening comprehension. In Y. N. Leung (Ed.), *elected Papers from the Twelfth International Symposium on English Teaching* (Vol. 2, pp. 64–72). Taipei, Taiwan: Crane.
- Field, J. (2003a). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57(4), 325–334.
- Field, J. (2003b). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57(4), 325–334.
- Flege, J. E. (1980). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics*, 5, 323–347.
- Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, 19, 407–439.
- Gilbert, J. B. (2001). *Clear speech*. Cambridge: Cambridge University Press.
- Gimson, A. C. (1989). *An introduction to the pronunciation of English* (4th ed.). Edward Arnold.
- Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55–75.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.



- Goodwin, J. M. (2001). Teaching Pronunciation. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 117–137). Boston, MA.: Heinle & Heinle.
- Grant, L. (2009). *Well said: Pronunciation for clear communication* (3rd ed.). Heinle & Heinle Pub.
- Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics*, 24(4), 495–522.
- Hardison, D. M. (2004). Generalization of computer-assisted prosody training: quantitative and qualitative findings. *Language Learning and Technology*, 8(1), 34–52.
- Hardison, D. M. (2005). Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal*, 22(2), 175–190.
- Hatch, E. M. (1992). *Discourse and language education*. Cambridge: Cambridge University Press.
- Heatley, A., Nation, P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>. Last accessed on September 11, 2013.
- Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments. *Psychological Methods*, 1, 154–169.
- Henrichsen, L. E. (1984). Sandhi-variation: a filter of input for learners of ESL. *Language Learning*, 34(3), 103–123.
- Herron, C. A., & Seay, I. (1991). The effect of authentic oral texts on student listening comprehension in the foreign language classroom. *Foreign Language Annals*, 24, 487–495.
- Hewings, M., & Goldstein, S. (1998). *Pronunciation plus student's book: Practice through interaction*. Cambridge University Press.
- Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech*, 27, 343–354.
- Hieke, A. E. (1987). Absorption and fluency in native and non-native casual speech in English. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition*. Dordrecht, The Netherlands; Providence, R.I.: Foris.
- Hieke, A. E. (1989). Spoken language phonotactics implications for the ESL/EFL classroom in speech production and perception. *Language Sciences*, 11(2), 197–213.
- Hirata, Y. (2004). Computer assisted pronunciation training for English native speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, 17(3-4), 357–376.

- Iowa State University Office of Institutional Research. (2013). *Iowa State University Fact Book 2012-2013*. Retrieved from <http://www.ir.iastate.edu/FB13>. Last accessed on September 23, 2013.
- Ito, Y. (2006). The comprehension of English reduced forms by second language learners and its effect on input-intake process. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 67–81). Manoa; Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- James, E. (1976). The acquisition of prosodic features of speech using a speech visualizer. *International Review of Applied Linguistics*, 14, 227–243.
- Johnson, K., Dunkel, P., & Rekart, D. (1994). Utilizing CALL to improve the English pronunciation of international teaching assistants. *CAELL Journal*, 5(2), 10–18.
- KayPENTAX. (2013). *KayPENTAX*. Retrieved from <http://www.kayelemetrics.com>. Last accessed on May 11, 2014.
- Kirsner, K. (1994). Implicit processes in second language learning. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 283–312). San Diego, CA: Academic Press.
- Kumai, N., & Timson, S. (2003). *Hit parade listening* (2nd ed.). Tokyo: MacMillan Language House.
- Kuo, H. C. (2009). *The effect of English linking instruction on EFL elementary school students' speech production and phonological awareness* (Unpublished MA Thesis). National Chung Cheng University, Chiayi, Taiwan.
- Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, 12(2), 137–156.
- Lambacher, S. (2001). The taming of English vowels. In *Proceedings of the PAC3/JALT 2001 Convention* (pp. 1–5). Japan.
- Lane, L. (2004). *Focus on pronunciation 3 (with 2 student audio CDs)* (2nd ed.). Pearson ESL.
- Lane, L., Mitchell, P., Molholt, G., Pennington, M., Perdreau, C., Cessaris, A., & Fischer, L. (1988). Taking the mystery out of the mouth: Invisible speech. Presented at the 22nd Annual TESOL Convention, Chicago, IL.
- Lara, M. Q. (2009). *The effects of acoustic visual feedback instruction on pronunciation of the English high front vowels by pre-service non-native English teachers* (Unpublished Dissertation). University of Kansas.
- Lass, R. (1984). *Phonology*. Cambridge: Cambridge University Press.

- Lee, J.-T. (2012). *A comparative study on the effectiveness of communicative and explicit connected speech instruction on Taiwanese EFL junior high school students' listening comprehension* (Unpublished MA Thesis). National Chunghua University of Education, Taiwan.
- Lee, J.-T., & Kuo, F.-L. (2010). Effects of teaching connected speech on listening comprehension. In *Selected Papers from the Nineteenth Symposium on English Teaching* (pp. 153–162).
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27.
- Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32, 505–524.
- Lin, H., Fan, C., & Chen, C. (1995). Teaching pronunciation in the learner-centered classroom. Presented at the 12th TEFL Conference, Tunghai University, R. O.C.
- Liu, J.-F. (2010). *Effects of using reduced-form keywords captions with graphical annotations on English learning* (Unpublished Thesis).
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1242–1255.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Routledge.
- Matsuzawa, T. (2006). Comprehension of English reduced forms by Japanese business people and the effectiveness of instruction. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 59–66). Manoa; Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McNerney, M., & Mendelsohn, D. (1992). Suprasegmentals in the pronunciation class: Setting priorities. In *Teaching American English pronunciation*. Boston, MA.: Houghton Mifflin.
- Melenca, M. A. (2001). *Teaching connected speech rules to Japanese speakers of English so as to avoid a staccato speech rhythm* (Unpublished Thesis). Concordia University.
- Molholt, G. (1988). Computer-assisted instruction for Chinese speakers of American English. *TESOL Quarterly*, 22, 91–111.
- Motohashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning and Technology*, 13(2), 29–47.

- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97.
- Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62, 715–719.
- Nation, P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Neri, A., Cucchiari, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5), 441–467.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.
- Norris, R. W. (1994). Keeping up with native speaker speed: An investigation of reduced forms and deletions in informal spoken English. *Studies in Comparative Culture*, 25, 72–79.
- O'Brien, M. (2006). Teaching pronunciation and intonation with computer technology. In Ducate, L. & Arnold, N. (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (pp. 127–148). San Marcos, Texas: Computer Assisted Language Instruction Consortium (CALICO).
- Omaggio Hadley, A. C. (1993). *Teaching language in context: Proficiency-oriented instruction* (2nd ed.). Boston, MA.: Heinle & Heinle.
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26–45.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45, 255–287.
- Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, 12(5), 427–440.
- Pennington, M. C., & Esling, J. H. (1996). Computer-assisted development of spoken language skills. In *The power of CALL* (pp. 153–189). Athelstan.
- Pennington, M. C., & Richards, J. C. (1986). Pronunciation Revisited. *TESOL Quarterly*, 20(2), 207–225.
- Pinker, S. (1995). *The language instinct: How the mind creates language*. New York: HarperPerennial.
- Pitch Instruments Inc. (2013). *WinPitch*. Retrieved from <http://www.winpitch.com>. Last accessed on May 15, 2014.

- Reed, M., & Michaud, C. (2005). *Sound concepts an integrated pronunciation course*. New York: McGraw-Hill.
- Rogers, C. V., & Medley, F. W. (1988). Language with a purpose: Using authentic materials in the foreign language classroom. *Foreign Language Annals*, 21, 467–478.
- Rogerson, M. (2006). Don'cha know? A survey of esl teachers' perspectives on reduced forms instruction. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 85–98). Manoa; Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- Rost, M. (2006). Areas of research that influence L2 listening instruction. In E. Usó Juan & A. Martínez Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 47–74). Berlin; New York: M. de Gruyter.
- Ruellot, V. M. (2006). *French pronunciation learning and computer-mediated visual feedback* (Unpublished Dissertation). University of Illinois at Urbana-Champaign.
- San, P. H. (2010). *The phonological development of Malaysian English speaking Chinese children: A normative study* (Unpublished doctoral dissertation). University of Canterbury, Christchurch, New Zealand.
- Sardegna, V. G. (2011). Pronunciation learning strategies that improve ESL learners' linking. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference* (pp. 105–121). Ames, IA: Iowa State University.
- Scheibman, J. (2000). I dunno: A usage-based account of the phonological reduction of don't in American English conversation. *Journal of Pragmatics*, 32, 105–124.
- Segalowitz, S. J., & Lane, K. C. (2000). Lexical access of function versus content words. *Brain and Language*, 75, 376–389.
- Sheeler, W. D., & Markley, R. W. (1991). *Sounds and rhythm* (2nd ed.). New Jersey: Prentice Hall.
- Shi, R., Gick, B., Kanwischer, D., & Wilson, I. (2005). Frequency and category factors in the reduction and assimilation of function words: EPG and acoustic measures. *Journal of Psycholinguistic Research*, 34(4), 341–364.
- Shockey, L. (1974). Phonetic and phonological properties of connected speech. *Ohio State Working Papers in Linguistics*, 17, 1–143.
- Shockey, L. (2003). *Sound patterns of spoken English*. Malden, MA: Blackwell Pub.
- Sinclair, J. (1991). *Corpus Concordance and Collocation*. Oxford: Oxford Univ Pr.
- Sinor, M. (2006). *Lexical and phonotactic cues to speech segmentation in a second language* (Unpublished Dissertation). University of Alberta.

- Skehan, P. (1998). *A Cognitive approach to language learning*. New York: Oxford University Press, USA.
- Spaai, W., & Hermes, D. (1992). A visual display for the teaching of intonation. *CALICO Journal*, 10(3), 19–30.
- Temperley, M. S. (1987). Linking and deletion in final consonant clusters. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory*. Teachers of English to Speakers of Other Languages.
- Tench, P. (1996). *The intonation systems of English*. London: Cassell.
- Ting, W.-Y., & Kuo, F.-L. (2012). Messages behind the unheard sounds: Crossing the word boundaries through songs. *NCUE Journal of Humanities*, 5, 75–92.
- Topping, D. (1964). Contrastive analysis and sandhi alterations. *Language Learning*, 14, 99–107.
- Trammell, R. L. (1999). English Ambisyllabic Consonants and Half-Closed Syllables in Language Teaching. *Language Learning*, 49, 311–356.
- Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24, 3–25.
- Vardanian, R. M. (1964). Teaching English intonation through oscilloscope displays. *Language Learning*, 14, 109–117.
- Varden, J. K. (2006). Visualizing English speech reductions using the free phonetic software package WASP. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 127–165). Manoa; Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa ; Distributed by University of Hawai'i Press.
- Verhofstadt, K. (2002). *A critical analysis of commercial computer-assisted pronunciation materials* (Unpublished doctoral dissertation). University of Ghent.
- Wang, Y. T. (2005). *An exploration of the effects of reduced forms instruction on EFL college students' listening comprehension* (Unpublished MA Thesis). National Tsing Hua University, Hsinchu, Taiwan.
- Watterson, B. (2008). *Calvin and Hobbes*. GoComics. Retrieved from <http://www.gocomics.com/calvinandhobbes/2008/08/14>. Last accessed on May 15, 2014.
- Weltens, B., & de Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech*, 27(1), 79–88.
- Wright, S. (1986). The interaction of sociolinguistic and phonetically-conditioned CPSs in Cambridge English: Auditory and electropalatographic evidence. *Cambridge Papers in Phonetics and Experimental Linguistics*, 5, (no page numbers).

- Wu, N. C. (1993). Contrastive study of Mandarin Chinese and American English phonological system. *Journal of Taipei Business Junior College*, 41, 361–465.
- Yang, J.-C., Lin, Y.-L., & Chung, C.-I. (2009). Developing a film-based learning system with English verbal reduced forms for supporting English listening comprehension. In M. Chang, R. Kuo, Kinshuk, G.-D. Chen, & M. Hirose (Eds.), *Learning by playing: Game-based education system design and development* (Vol. 5670, pp. 74–84). Berlin, Heidelberg: Springer Berlin Heidelberg.

## APPENDIX I. PRE-TRAINING QUESTIONNAIRE

### Teaching Connected Speech Features to Non-Native Speakers of English

#### I. Background information

1. Name .....
2. Nickname (to be used in the study) .....
3. Home country ..... 4. Native language .....
5. Gender            Male            Female
6. Age .....
7. Degrees/Major .....
8. Do you have any speech, hearing, or visual impediment (problem)? Yes \_\_\_\_\_ No \_\_\_\_\_
9. If yes, what is it?  
.....

#### II. Language Study

1. How long have you been living in the U.S.? .....
2. How long did you study English for in your home country? .....
3. What kind of pronunciation instruction did you receive before you came to the U.S. (if any)  
.....  
.....  
.....



4. What kind of pronunciation instruction have you received in the U.S. (if any)

.....  
 .....

5. What is your most recent TOEFL score, if applicable .....

6. What is your most recent OECT score (Speak/Teach), if applicable

.....

7. Spoken American English is difficult to understand.

Agree

Somewhat agree

Somewhat disagree

Disagree

8. What difficulties do you find in understanding English pronunciation?

9. Have you noticed how English speakers connect their words to one another?

Yes

No

Explain what you noticed:

10. Have you received any teaching/training about linking words in pronunciation and/or pausing?

Yes

No

Please explain:

## APPENDIX II. INFORMED CONSENT DOCUMENT

---

**Title of study:** Teaching Connected Speech Features to Non-Native Speakers of English

**Investigator:** Ghinwa Alameen

---

I would like to invite you to participate in this study and ask that you read this document and ask any questions you have before agreeing to be in the study.

My name is Ghinwa Alameen, and I am a Ph. D. Student in Applied Linguistics and Technology at Iowa State University. The purpose of this study is to examine features of connected speech of non-native speakers of English. It will also examine whether computer technology can improve learners' pronunciation of such features. These findings will help shape and focus our efforts in teaching English speaking skills. You are being invited to participate in this study because you are a non-native learner of English.

If you agree to participate, you will do the following:

- 1) Take a pretest, a posttest and a delayed posttest. The pretest will take place before the training, the posttest after the training, and the delayed posttest one month after the training. For each of these tests, you will be asked to (a) write down recorded English sentences, (b) read a group of English sentences. Your speech will be recorded for analysis. Each test should take about 30 minutes.
- 2) May participate in pronunciation training using different types of computer technologies. Training will take place over a period of two weeks for 6 sessions. Each session will last 40 minutes.
- 3) Fill out (a) a background questionnaire that gathers information about your age, nationality, language proficiency, etc., and (b) a questionnaire about your views of the effectiveness of the training. The questionnaires should take about 30 minutes in total. Please note that you do not have to answer all questions of the questionnaires in order to participate.

This study does not have any risks. There are no direct benefits of being in the study. It may help you improve your pronunciation. In addition, the results may help teachers develop more efficient ways to teach speaking skills to non-native speakers of English.

Your participation in this study is voluntary, and you may refuse to participate or stop participating at any time once the study has started. If you decide to not participate in the study or leave the study early, it will not result in any penalty.

The records of this study will be kept confidential. Your identity will be protected: upon your participation, you will choose a nickname that will be used with your data. All materials will be entered and saved on my personal computer and protected by a password. I am the only person who will have access to the computer and the passwords. All materials and data will be deleted and destroyed once the research is complete. If the results are published, your identity will remain confidential. Records identifying participants will be kept confidential to the extent allowed by applicable laws and regulations. Records will not be made publicly available. However, federal government regulatory agencies, auditing departments of Iowa State University, and the ISU Institutional Review Board (a committee that reviews and approves research studies with human subjects) may inspect and/or copy your records for quality assurance and analysis. These records may contain private information.

If you do not understand any portion of what you are being asked to do, or the contents of this form, I am available to provide a complete explanation. Your questions are welcome at any time. Please do not hesitate to ask your questions and/or mention your concerns during our meetings. Feel free to contact me, Ghinwa Alameen, [ghinwa@iastate.edu](mailto:ghinwa@iastate.edu). You can also contact Dr. John Levis, the supervisor of this research, 515-294-7524, [jlevis@iastate.edu](mailto:jlevis@iastate.edu). If you have any questions about the rights of research subjects, please contact the IRB (Human Subjects) Administrator, 515-294-4566, [IRB@iastate.edu](mailto:IRB@iastate.edu), or Director, Office for Responsible Research, 515-294-3115, 1138 Pearson Hall, Ames, IA 50011.

---

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document, and that your questions have been answered satisfactorily.

Participant's Name (printed) \_\_\_\_\_

\_\_\_\_\_

(Participant's Signature)

\_\_\_\_\_

(Date)

I certify that the participant has been given adequate time to read and learn about the study and all of his or her questions have been answered. It is my opinion that the participant understands the purpose, risks, and benefits, and the procedures in this study, and has agreed voluntarily to participate.

\_\_\_\_\_

(Signature of Person Obtaining Informed Consent)

\_\_\_\_\_

(Date)

## APPENDIX III. AV GROUP TRAINING HANDOUT

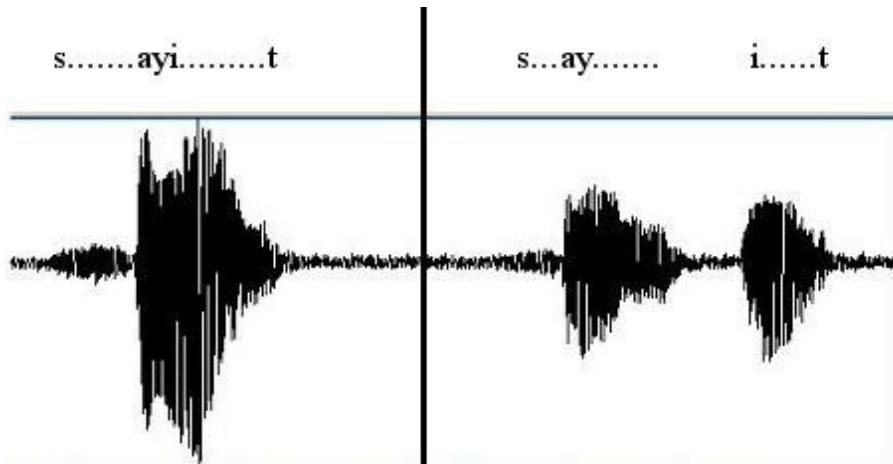
### Instructions for Day 2 - AV

Welcome to the Connected Speech Training!

The goal of these sessions is to help you improve your ability to listen to and pronounce connected speech. This handout will give you instructions on how to record yourself using Audacity and how to understand the waveforms provided on it.

1. Go to 'Applications' and find Audacity.
2. To record, just click the red Record button. When done, quickly click the Stop button.
3. Log in to the training website
4. Go to "Help with recording and waveforms". Every screen will have a sentence/phrase, an audio and a waveform.
5. Read the first sentence '*say it*' a couple of times for yourself. Try to connect/link the two words together. Then raise your eyes and record the sentence on Audacity. Try not to read while recording and be natural. Audacity will produce a waveform of your recording.
6. Now go back to our training website. Listen to the sentence '*say it*' there and repeat it after the speaker. Compare your waveform to the waveform on the website. The two waveforms should be similar but not exactly the same. They should follow the same pattern of thick and thin waves.
7. If the two waveforms are not similar, then listen again to the native speaker audio and keep recording until your waveform is similar to the one on the website.
8. Make sure to compare the two waveforms closely to find which part(s) is causing the difference.

9. For example, Look at the following two pronunciations of ‘say it’. The one on the left is linked, so the waveforms of the two words ‘say’ and ‘it’ are very close to each other. The one on the right is NOT linked so the two waveforms have separate waveforms with a little silence (pause) in between them. Your sentence should look like the one on the left.

Linked ‘say it’

Not linked ‘say it’

10. Please note that the waveform of linked/connected words is not also combined (and fat) like the above one. Sometimes there is a thin wave that happens because of there is a voiceless consonant in the word like /s, t, p, k, f/.

## Overview of recording instructions

1. Read the sentence on your screen several times



2. Remember to LINK the words



3. Raise your eyes and RECORD on Audacity



4. Listen to the audio online  compare with your recording



5. Compare your waveform with the one online



6. Record again till the two waveforms have a similar pattern

Done



**APPENDIX IV. PERCEPTION TEST WORKSHEET**

<b>Pronunciation Worksheet</b>	Name: _____
--------------------------------	-------------

Listen to the recording and write down the sentences you hear on the worksheet. You can only listen once. If you are not sure of the words, write your best guess.

1. ....

2. ....

3. ....

4. ....

5. ....

6. ....

7. ....

8. ....

9. ....

10. ....

11. ....

12. ....

13. ....



## **APPENDIX V. PERCEPTION TEST STIMULI**

### **Perception Pretest**

1. Ames is a city in the west.
2. Play it again.
3. She puts on her makeup.
4. Come over to my house.
5. View our latest collection.
6. Come up with an answer.
7. They have a dozen apples.
8. We don't see him too often.
9. You'll do all of the work.
10. I'll see her in an hour.
11. I had an apple and a cup of orange juice.
12. That's an easy example.
13. He looked at everything around him.

**Perception Posttest**

1. That's an easy example.
2. He likes to go out with a friend.
3. Tell her I'll be there.
4. Let's go over our notes.
5. View our latest collection.
6. We don't see him too often.
7. They have a dozen apples.
8. Come up with an answer.
9. She puts on her makeup.
10. Ames is a city in the west.
11. Come over to my house.
12. Play it again.
13. He looked at everything around him.
14. It turned out to be a wonderful day.
15. I'll see her in an hour.
16. I had an apple and a cup of orange juice.
17. Is he in his office?
18. We ran out of milk.
19. It fell out of the box.
20. Some people are open to this idea.
21. Just do it right away.
22. I know it's a little dirty inside.
23. Nobody pays attention to it.
24. She came in time for dinner.
25. Can you help us with our kitchen?
26. You'll do all of the work.

## APPENDIX VI. PERCEPTION TEST STIMULI LINKING

Pretest	Posttest New Links
<p>Ames is is a city in play it it again puts on on her come over view our come up with an an answer have a dozen apples see him too often do all all of see her her in in an an hour had an an apple apple and and a cup of of orange that's an an easy easy example looked at at everything everything around around him</p>	<p>people are are open this idea turned out be a fell out out of is he he in in his his office ran out out of tell her her I go over over our do it right away go out with a pays attention to it help us with our came in know it's it's a dirty inside</p>

## APPENDIX VII. PRODUCTION TEST STIMULI

### Production Pretest

#### Part One

A little boy was walking down a road, when he came across a snake. /  
 The snake asked / "Take me to the top of the mountain please. / I hope to see the  
 sunset before I die." / The little boy answered "No. / If I pick you up, you will  
 bite me and I will die." /

The snake asked him not to be afraid. / He agreed and carried it up  
 to the top of the mountain. / They sat under a tree and watched the sunset. /

Then after sunset, / the snake asked / "Can I go home now. / I'm tired and I'm old."

#### Part Two

You may have thought that foreign language is just ambiguous speech, / but it isn't. /  
 Children have an intrinsic ability to acquire languages. / They can implicitly pick up new  
 words after they're exposed to them once or twice. /

They seem to implement an array of strategies that adults do not have access to. /

However, / as they grow up, / they encounter more difficulties in obtaining new  
 words and utilizing such skills. /

## Production Posttest

### Part One

A little boy was walking down a road, when he came across a snake. /

The snake asked / "Take me to the top of the mountain please. / I hope to see the sunset before I die." / The little boy answered "No. / If I pick you up, you will bite me and I will die." /

The snake asked him not to be afraid. / He agreed and carried it up to the top of the mountain. / They sat under a tree and watched the sunset. /

Then after sunset, / the snake asked / "Can I go home now. / I'm tired and I'm old."

The little boy felt he was safe all this time with the snake, so he would take it home. / He picked it up and carried it back to its home. /

Before he put it on the ground, / it turned and bit him. / The boy cried out and threw it away from him. / He asked / "Mr. Snake, why did you do that? Now I will surely die." /

The snake looked up at him and answered / "You knew what I was when you picked me up." /

---

## Part Two

You may have thought that foreign language is just ambiguous speech, / but it isn't. /  
 Children have an intrinsic ability to acquire languages. / They can implicitly pick up new  
 words after they're exposed to them once or twice. /

They seem to implement an array of strategies that adults do not have access to. /

However, / as they grow up, / they encounter more difficulties in obtaining new  
 words and utilizing such skills. /

Their bodies undergo many adjustments. / Their voices are altered. /  
 Many of them are not aware of those internal changes. /

They only rarely invoke images of the children they used to be. /

In order to access their language skills / adults should adapt to a variety of conditions. /  
 Using such tools will not only help them accommodate to new situations, /  
 but also be able to indicate sources of confusion. /

## APPENDIX VIII. LINKING IN PRODUCTION TEST STIMULI

	Pretest HF text	Novel HF text	Pretest LF text	Novel LF text
<b>Linking pairs</b>	down a when he came across across a snake asked top of before I boy answered If I you up me and and I snake asked asked him be afraid he agreed carry it it up top of sat under under a tree and then after snake asked can I tired and and I	take it picked it it up carried it to its put it it on cried out threw it it away he asked now I looked up up at at him him and and answered what I me up	language is just ambiguous but it it isn't have an an intrinsic intrinsic ability to acquire can implicitly pick up are exposed once or implement an an array array of that adults have access grow up they encounter difficulties in in obtaining words and	bodies undergo many adjustments voices are are altered many of them are not aware aware of those internal they only images of in order should adapt to a variety of not only them accommodate but also be able to indicate sources of
<b>Discarded linking pairs</b>	am old	felt he safe all so he up and before he turned and bit him out and from him	words after to implement	rarely invoke invoke images to access

## APPENDIX IX. PERCEPTION ITEM ANALYSIS

Items	AO			AV			Combined		
	Pretest	Posttest	Difference	Pretest	Posttest	Difference	Pretest	Posttest	Difference
Ames is	100	100	0	87	93	7	93	97	3
is a	67	53	-13	60	60	0	63	57	-7
city in	80	73	-7	87	87	0	83	80	-3
play it	87	93	7	80	100	20	83	97	13
it again	87	93	7	80	100	20	83	97	13
puts on	93	93	0	93	93	0	93	93	0
on her	40	67	27	60	73	13	50	70	20
come over	93	80	-13	87	100	13	90	90	0
view our	20	40	20	13	33	20	17	37	20
come up	93	100	7	87	100	13	90	100	10
with an	33	40	7	27	47	20	30	43	13
an answer	33	40	7	27	47	20	30	43	13
have a	60	87	27	60	80	20	60	83	23
dozen apples	7	40	33	40	40	0	23	40	17
see him	40	60	20	40	47	7	40	53	13
too often	80	93	13	93	93	0	87	93	7
do all	93	87	-7	93	93	0	93	90	-3
all of	73	67	-7	73	87	13	73	77	3
see her	0	0	0	7	7	0	3	3	0
her in	7	0	-7	40	7	-33	23	3	-20



in an	80	93	13	67	93	27	73	93	20
an hour	93	93	0	67	93	27	80	93	13
had an	80	87	7	67	93	27	73	90	17
an apple	80	87	7	60	87	27	70	87	17
apple and	80	87	7	60	87	27	70	87	17
and a	47	67	20	60	80	20	53	73	20
cup of	40	40	0	53	47	-7	47	43	-3
of orange	67	67	0	93	80	-13	80	73	-7
that's an	67	67	0	60	73	13	63	70	7
an easy	67	73	7	53	73	20	60	73	13
easy example	100	100	0	100	100	0	100	100	0
looked at	20	20	0	33	33	0	27	27	0
at everything	20	20	0	33	33	0	27	27	0
everything around	80	93	13	87	100	13	83	97	13
around him	73	80	7	87	93	7	80	87	7

## APPENDIX X. PRODUCTION ITEM ANALYSIS

Items	AO			AV			Combined			NSs
	Pretest	Posttest	Difference	Pretest	Posttest	Difference	Pretest	Posttest	Difference	
down a	73	100	27	60	73	13	67	87	20	100
when he	0	13	13	7	0	-7	3	7	3	80
came across	40	67	27	53	87	33	47	77	30	100
across a	47	80	33	53	80	27	50	80	30	100
snake asked	20	27	7	7	20	13	13	23	10	60
top of	40	73	33	47	93	47	43	83	40	100
before I	73	93	20	67	87	20	70	90	20	100
boy answered	7	7	0	13	20	7	10	13	3	60
If I	73	93	20	60	73	13	67	83	17	100
you up	40	73	33	27	60	33	33	67	33	100
me and	7	13	7	0	13	13	3	13	10	80
and I	13	47	33	27	27	0	20	37	17	80
snake asked	13	20	7	7	13	7	10	17	7	40
asked him	0	27	27	0	0	0	0	13	13	60
be afraid	60	80	20	53	93	40	57	87	30	80
he agreed	33	67	33	67	73	7	50	70	20	80
carry it	33	53	20	47	67	20	40	60	20	80
it up	40	67	27	27	87	60	33	77	43	100
top of	53	87	33	60	73	13	57	80	23	100
sat under	13	33	20	7	13	7	10	23	13	100
under a	27	73	47	27	73	47	27	73	47	80

tree and	13	33	20	7	13	7	10	23	13	100
then after	20	27	7	7	7	0	13	17	3	40
snake asked	20	27	7	7	40	33	13	33	20	80
can I	93	100	7	100	93	-7	97	97	0	60
tired and	7	47	40	0	40	40	3	43	40	100
and I	27	73	47	33	53	20	30	63	33	60

## APPENDIX XI. POST-TRAINING QUESTIONNAIRES

(AV Group)

### Teaching Connected Speech Features to Non-Native Speakers of English

**I would ask you to complete the following questionnaire in order to help me improve the pronunciation training sessions for future students. Please provide as much feedback as possible. Thank you for your participation!**

**I. Please check the answer that applies to you best and write any comments you have about it.**

1. Overall, I liked the past pronunciation training sessions.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

2. Being able to record (and re-record) myself while practicing helped me improve my pronunciation.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

3. I found it helpful to compare the waveform of the model speaker with my own.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

4. The explanation of how to interpret waveforms helped me improve the way I use them.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

5. I would like more instruction similar to this training.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

6. The waveform activity I worked on was helpful for noticing how words are connected.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

7. I can recognize connected words better in speech after this training.

	Agree	Somewhat agree	Somewhat disagree	Disagree
8. This training helped me notice how I connect words.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
9. It was difficult to see the link between words in the waveform.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
10. It was difficult to create a waveform.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
11. After the training on electronic visual feedback, I improved the pronunciation of the English connected speech.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
12. Electronic visual feedback allows me to work independently on my pronunciation.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
13. Electronic visual feedback allows me to self-monitor my pronunciation.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
14. I understood all the vocabulary words used in each practice session.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
15. The number of practice sessions was appropriate.				
	Agree	Somewhat agree	Somewhat disagree	Disagree
16. The length of practice sessions was appropriate.				
	Agree	Somewhat agree	Somewhat disagree	Disagree

**II. Please answer the following questions and provide as much feedback as possible.**

1. What have you noticed about your own pronunciation in English as a result of this training?
2. If this training was confusing or difficult, please explain why.
3. What did you like the **most** about the sessions?
4. What did you like the **least** about the sessions?
5. What modifications would you suggest to improve these training sessions?

## (AO Group Post-training Questionnaire)

**Teaching Connected Speech Features to Non-Native Speakers of English**

**I would ask you to complete the following questionnaire in order to help me improve the pronunciation training sessions for future students. Please provide as much feedback as possible. Thank you for your participation!**

**I. Please check the answer that applies to you best and write any comments you have about it.**

1. Overall, I liked the past pronunciation training sessions.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

2. I can understand connected words better in speech after this training.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

3. I would like more training similar to this training.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

4. This training helped me notice how I connect words.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

5. It was easy to compare my recording to the audio on the website.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

6. The number of practice sessions was appropriate.

Agree                      Somewhat agree                      Somewhat disagree                      Disagree

7. Being able to record (and re-record) myself while practicing helped me improve my pronunciation.

Agree	Somewhat agree	Somewhat disagree	Disagree
-------	----------------	-------------------	----------

8. The teacher's explanation of connected speech helped me understand it better.

Agree	Somewhat agree	Somewhat disagree	Disagree
-------	----------------	-------------------	----------

9. This online practice allows me to work independently on my listening and pronunciation.

Agree	Somewhat agree	Somewhat disagree	Disagree
-------	----------------	-------------------	----------

10. The examples taken from movies were easy to understand.

Agree	Somewhat agree	Somewhat disagree	Disagree
-------	----------------	-------------------	----------

11. I'd like to have more examples taken from movies.

Agree	Somewhat agree	Somewhat disagree	Disagree
-------	----------------	-------------------	----------

## **II. Please answer the following questions and provide as much feedback as possible.**

1. What have you noticed about your own pronunciation in English as a result of this training?

2. If this training was confusing or difficult, please explain why.

3. What did you like the **most** about the sessions?

4. What did you like the **least** about the sessions?

5. What modifications would you suggest to improve these training sessions?